Blueprint Labs

Discussion Paper #2025.03

The Impact of Increased Exposure of Diversity on Suburban Students' Outcomes: An Analysis of the METCO Voluntary Desegregation Program

Elizabeth Setren

March 2025

MIT Department of Economics
77 Massachusetts Avenue, Bldg. E53-390
Cambridge, MA 02139

National Bureau of Economic Research
1050 Massachusetts Avenue, 3rd Floor
Cambridge, MA 02138

The Impact of Increased Exposure of Diversity on Suburban Students' Outcomes: An Analysis of the METCO Voluntary Desegregation Program

Elizabeth Setren*

March 20, 2025

Abstract

Over sixty years following Brown vs. Board of Education, racial and socioeconomic segregation and lack of equal access to educational opportunities persist. Across the country, voluntary desegregation busing programs aim to ameliorate these imbalances and disparities. A longstanding Massachusetts program, METCO, buses K-12 students of color from Boston and Springfield, Massachusetts to 37 suburban districts that voluntarily enroll urban students. Supporters of the program argue that it prepares students to be active citizens in our multicultural society. Opponents question the value of the program and worry it may have a negative impact on suburban student outcomes. I estimate the causal effect of exposure to diversity through the METCO program by using two types of variation: difference-in-difference analysis of schools stopping and starting their METCO enrollment and two-stage least squares analysis of space availability for METCO students. Both methods rule out substantial test score, attendance, or suspension effects of having METCO peers. Classroom ability distribution and classroom suspension rates remain similar when METCO programs start and stop. There is no negative impact on college preparation, competitiveness, persistence, or graduation.

*Tufts University, Department of Economics, Medford, MA 02155, elizabeth.setren@tufts.edu. Thanks to seminar participants at CESifo Education Program, AEFP 2021, Syracuse University, Vanderbilt University, Harvard Program on Education Policy and Governance, Harvard Kennedy School, and Tufts University for helpful feedback. I am grateful to The Boston Foundation, the Institute for Education Sciences (Grant R305A200060), The Spencer Foundation, and The Russell Sage Foundation for providing generous financial support for this project. Many thanks to Savannah Kochinke and Elizabeth Pancotti for providing superb research assistance. Thanks also to Eryn Heying for administrative support. Special thanks go to Carrie Conaway, Cliff Chuang, Elana McDermott, Matthew Deninger, and the staff of the Massachusetts Department of Elementary and Secondary Education for data, suggestions, and assistance. I am incredibly grateful to Milly Arbaje-Thomas, Charles Walker, Jean McGuire, John Shandorf, Wilmary Tejeda, Kim Houston, Deborah Ward, Colin Stokes, Kristen Fumarola, Jeffrey Welch, and the rest of the METCO staff for sharing their expertise, feedback, historical records, and data.

In 2020, nationwide protests following the murder of George Floyd led to increased attention to racial equity in education. School integration received renewed focus as a potential policy response. Some of the nation's largest school districts, including New York City, Los Angeles, Miami-Dade, Dallas, Milwaukee, Philadelphia, and Oakland, formed a collaborative effort to increase racial and economic integration in their schools (Belsha and Darville, 2020). President Biden launched the Fostering Diverse Schools Demonstration Grants Program to fund school integration efforts (Office of Elementary & Secondary Education, 2023; Belsha, 2023). Even before these efforts, there was an increase in integration policies: 45 districts started integration programs in 2017-2020 and over 14 percent of public school students attend schools with active integration programs (Potter and Burris, 2020). These programs aim to reverse school segregation, which has increased in the past 25 years. Since 2001 the proportion of K-12 public schools with over 75 percent of poor students of color has doubled (Government Accountability Office, 2016). Research links school segregation to widening disparities in academic achievement (Card and Rothstein, 2007; Vigdor and Ludwig, 2007).

Within district integration efforts, such as redrawing school assignment boundaries, creating specialized schools, district-wide school choice, and transfer policies that prioritize low-income students are limited in their ability to increase integration since segregation most stems from racial disparities between school districts (Clotfelter, 1999; Reardon, Yun and Eitle, 2000; Logan, Stults and Farley, 2004; Logan, Oakley and Stowell, 2008). Across district integration could address the main source of segregation, but the Supreme Court ruled mandatory busing across district lines unconstitutional in Milliken v. Bradley (1974). This left voluntary across district integration programs as the key education policy tool to promote school integration.

Voluntary desegregation busing sends students of color from urban school districts to predominantly white, suburban schools that elect to participate.[1] A longstanding desegregation busing program, the Metropolitan Council for Educational Opportunity (METCO) serves as a national model for other desegregation programs throughout the country (Eaton, 2001). Founded in 1966, METCO buses students of color from Boston and Springfield, Massachusetts to 37 suburban K-12 school districts that voluntarily elect to enroll urban students. Currently, over 3,150 Boston and 150 Springfield students attend suburban districts through the METCO program. Since 2020, METCO has received requests to increase enrollment in participating districts and to expand to nine new suburban districts (Martin, 2021). The state-funded program aims to promote diversity and cultural competency by reducing racial isolation in suburban and urban districts and to increase access to high performing schools for urban students.

Supporters of efforts to increase school diversity (like METCO) assert that integration has positive peer effects by exposing students to different cultures, backgrounds and views. Research suggests reduced racial

[1] East Palo Alto, Omaha, St. Louis, Minneapolis, Milwaukee, New Haven, Hartford, and Rochester each have voluntary desegregation busing programs.

isolation may help students overcome racial stereotypes, learn from other cultures, and reduce prejudice (Godsil et al., 2014; Zebrowitz, White and Wieneke, 2008). Others express concern that there may be negative peer effects on academics and behavior from increased racial and socioeconomic diversity (McGuirk, 2019; Town of Brookline Board of Selectmen, 2014; Cohen, 1990; Vigue, 1999). Opposition to school integration efforts still results in protests and resistance as seen in Howard County, Maryland (Goldstein, 2019). More generally, parents' perception of school quality and reputation is highly influenced by the share of students of color (Wells, 2015; Ellen, 2000). These concerns, whether implicit or explicit, counteract integration efforts and worsen school segregation.

This paper provides causal estimates on whether increased exposure of urban students of color has academic or behavioral peer effects on suburban students using two different identification strategies with different local average treatment effects. The first method uses difference-in-differences analysis to compare cohorts with METCO students to cohorts in the same school without METCO students. This identifies the effect of having any METCO peers in the grade. The second method harnesses the fact that classroom space constraints play a role in determining the number of METCO students a school accepts each year in a two-stage least squares (2SLS) analysis similar to Angrist and Lang (2004)'s adaptation of Maimonides' rule. This strategy identifies the effect of an additional METCO student per classroom.

This paper contributes to the literature on the impacts of school integration and school choice policies that increase integration. Court-ordered integration generated strong positive effects for Black students (Guryan, 2004; Johnson, 2015; Tuttle, 2019; Reber, 2010; Ashenfelter, 2006) and removal of integration orders had detrimental effects (Billings, Deming and Rockoff, 2014; Gamoran and An, 2016; Lutz, 2011; Saatcioglu, 2010). Voluntary integration generates large gains in college going and graduation for students of color in METCO (Setren, 2024) and mixed effects in a California program (Bergman, 2018). Court-ordered desegregation led white students to leave for other schools (Baum-Snow and Lutz, 2011; Liebowitz and Page, 2014; Reber, 2005; Welch and Light, 1987), but had no impact on white students' educational attainment (Johnson, 2015; Guryan, 2004).

Evidence on the effects of voluntary integration on white and higher income students are more limited and have mixed results. Domina et al. (2021) and Hill et al. (2023) find that a school assignment policy intended to reduce income segregation increased Math and English test scores, reduced suspension rates, but may have reduced achievement for students from higher-income backgrounds. Cook (2024) finds that switching from a voluntary racial integration program to race-blind admissions increased segregation and reduced student achievement and college enrollment for Black students and found non-persistent achievement losses for non-Black students. Angrist and Lang (2004) study one METCO suburban district from 1994 - 2000 and find no impact of increased diversity on suburban white students and suggestive evidence of a negative effect for Black suburban students, particularly girls.

This study's setting has several features that make it ideal for building upon existing literature and understanding the impact of current school integration efforts. The longevity and large size of the program allow for analysis of 38 different school districts and two urban areas (Boston and Springfield, Massachusetts) across two decades. Using recent data adds to the literature since the impact of integration may change as residents of suburban school districts become more diverse and state standardized testing has become higher stakes. Looking at both a large and mid-sized urban area and different school districts across time provides an understanding of the impact of integration in different settings and time periods. Early adopters of METCO or districts that accept relatively more METCO students (such as the district in Angrist and Lang (2004) which helped start the program and has one of the largest enrollments) may have different effects than later adopters or those that accept fewer METCO students. The degree of integration with urban students of color varies between 1 percent and 10 percent of the grade cohort which is a common range of increased integration from open enrollment and district choice programs. Similarly, school practices to support and integrate METCO students vary across districts and these may affect the impacts of the program. Having a variety of districts enables us to see the impact of increased integration in a variety of circumstances and explore heterogeneous effects. Detailed administrative data enables estimation of the impact on high school and college outcomes and analysis of how increased integration impacts classroom and teacher characteristics. Lastly, this paper also adds to the literature by exploring both the effect of any urban peers and the effect of an additional urban peer.[2]

I find no substantial negative impact on test scores, attendance, or suspensions from having METCO peers in the cohort or having an additional METCO student per class. Opponents to increased integration may be concerned that it could widen the skill distribution within a classroom which could worsen the match between the classroom content and students' ability level. I find that having METCO peers does not substantially widen the 90th - 10th percentile ability distribution in Math or English classes. Others may be concerned that increased integration would lead to increased classroom disruptions, but I find no substantial changes in the classroom suspension or attendance rates. There are also no negative impacts on college preparation outcomes or competitive college enrollment. Effects are similar across district and student characteristics. The conclusive null effects of urban students on the suburban students' outcomes

[2]This paper builds upon Angrist and Lang (2004) in several key ways. Angrist and Lang (2004) analyze the impact of having one additional METCO student in the classroom on peers' test scores. This paper conducts similar analysis and also estimates the effect of having any METCO students in the cohort on peers. Angrist and Lang (2004) analyze one district, Brookline, in the 1990s. Brookline was one of the founding districts of METCO, has among the largest METCO enrollment, and has more mixed income housing that other METCO suburban districts. Since the 1990s, emphasis on the state standardized test has increased substantially and expanded to other grade-levels. School districts have also become more diverse. This paper analyzes the impact on all 38 METCO districts which vary in terms of racial and economic diversity within the district and how many METCO students they accept. I explore heterogeneity, longer run outcomes, and behavior outcomes. I also can test potential mechanisms with classroom and teacher characteristic data.

suggests that the concerns of negative academic or behavioral effects from increases urban students of color in Boston and Springfield-area suburban schools are unwarranted.

1 Background

Origin of the METCO Program

METCO formed in response to a contentious battle over school integration in Boston Public Schools. In 1965, a Massachusetts State Board of Education study reported intense racial imbalance in Boston schools and concluded that it harmed both Black and white students. Black parents and students protested the inadequate school facilities and resources, overcrowding, and segregation in the Boston schools. The report and protests contributed to the passing of the Racial Imbalance Act of 1965 which defined racial imbalance as any school comprised of more than 50 percent non-white students and required the school to desegregate or risk losing their state funding. A resulting report "Because it is Right - Educationally" found that 44 of the 55 racially imbalanced Massachusetts schools were located in Boston (Advisory Committee on Racial Imbalance and Education, 1965). The State Board of Education instructed the Boston School Committee to develop a busing plan to integrate Boston Public Schools and then to implement it. The State Board of Education found the Boston School Committee's integration plan inadequate and voted to rescind state education funding to Boston in 1966. The Boston School Committee challenged this decision which brought the battle over integration to the courts. Ultimately Massachusetts Judge W. Arther Garrity, Jr. ruled in 1974 that the Boston's school enrollment and transfer policies enabled racial discrimination and he ordered desegregation busing. The ruling led to intense backlash with protests, riots, and violence across the city (Levy, 1971; Formisano, 1991).

In the wake of the Boston School Committee's opposition to school integration, Black parents and activists forged their own integration programs. In 1965, Operation Exodus utilized Boston's Open Enrollment Policy to bus over 400 students from a predominantly Black neighborhood in Roxbury to predominantly white schools around Boston with open seats. After the first year, METCO formed when nearby suburban districts' school committees agreed to accept Black students from Boston to fill empty seats and support school integration. In the Fall of 1966, 220 Kindergarten through 11th grade students from the city of Boston enrolled in Brookline, Lexington, Newton, Wellesley, Braintree, Arlington, Lincoln, and Concord. Operation Exodus ended in 1969 due to insufficient funds. The remaining suburban busing program, METCO, was intended to be temporary until Boston Public Schools integrated, but turmoil over integration in Boston persisted through the 1980s. During this time, the METCO program expanded to new suburban districts and increased enrollment within the original districts (Batson and Hayden, 1987).

A two-step process determined whether new suburban districts could join METCO. First, the town needed to approve through either a school committee or town council vote. Then METCO chose whether to accept the school district. In cases where towns had substantial opposition to accepting METCO students or only approved joining METCO by a small margin, METCO turned down districts.[3] New districts joined until 1975 and no new districts have been approved since. Towns have approached METCO to join the program and there have been nine new requests since 2020, but none have been approved. Milton, Rockland, Hamilton-Wenham, and Framingham ended their METCO programs, but the other 33 districts have continuously enrolled METCO students since they joined.[4] The city of Springfield also started its own smaller METCO program with four nearby suburbs. Springfield currently enrolls 150 K-12 students a year. Appendix Figure 1 shows the growth in METCO enrollment as new districts joined and relatively stable enrollment of over 3,000 students since the 1980s. The participating suburbs included two that directly bordered Boston (Newton and Brookline) and others over 20 miles away from Boston (see Appendix Figure 5 for a map of the participating districts). The farthest district is approximately 30 miles from Boston's center.

METCO Program Enrollment & Services

Decisions of whether and how many METCO students to enroll are made by district and school leaders in the Spring. These decisions could be influenced by the projected space in classrooms, budget changes and concerns, efforts to have a fixed number of total METCO students in the district across all grades, and idiosyncratic preferences of school and district leadership. Over 90.9 percent of METCO-receiving districts accept METCO students in either Kindergarten or first grade. Once a suburban district enrolls a METCO participant from Boston or Springfield, they commit to educating them through the 12th grade as long as the student continues to live in the city. As such, most of the cohort-level variation in METCO exposure starts in Kindergarten or first grade and continues throughout primary and secondary school. Four districts started METCO enrollment in middle or high school during the study period,[5] but otherwise a relatively small number of students start in the program after the entry grades. After the schools in a district decide on the number of METCO students that they want to enroll, then that number of applicants from the METCO waitlist are connected with the district to enroll. Details about the application and selection process are discussed in Setren (2024).

[3] METCO rejected Winchester in 1967 and 1974. METCO also rejected Beverly, Georgetown, and Randolph in 1974.

[4] Framingham ended their METCO participation during the study period so their program contributes to the analysis sample. The three other programs ended before the study period.

[5] Natick, Westwood, Foxborough, and Framingham had middle or high school entry.

Each suburban district has a METCO Director who oversees the program and supports the METCO participants. Suburban districts vary in the degree of additional supports they provide their urban students. Appendix Table 1 shows the types and prevalence of supports across the districts. These can include late buses, so that students can participate in after school activities, tutoring, access to social workers and counselors, and social programs to foster relationships with students and families at the school. In addition, METCO urged the suburban districts to have inclusive curriculums that are free from racial and gender stereotypes and include African and African-American history and culture in all disciplines.[6] METCO also encouraged the districts to hire Black teachers and staff.[7]

Reasons for Suburban Support and Opposition

Districts joined METCO to increase the diversity of their schools and prepare their students to interact with people from different backgrounds. Both METCO and suburban districts cited this reason, saying it "would broaden the experience"[8] of students and create "a new learning experience for suburban children."[9] Moral responsibility and charity also motivated some suburbanites to participate. METCO pushed back against this reason because it ignored the benefit urban students brought to the suburban schools and stereotyped METCO students as all poor (Tutino, 1966). While the diversity in the suburban schools has increased since they joined the METCO program in the 1960s and 1970s, they still remain predominantly white and the motivation to increase school diversity remains today. There is policy discussion around expanding METCO enrollment and eight new districts have inquired about joining the program.[10]

Opponents to the METCO program focus on cost concerns and potential negative spillover effects on suburban students. The state of Massachusetts funds the METCO program (the next section provides details about funding). Residents expressed concern that property taxes would increase to pay for the program and that they didn't consent to have their state tax dollars spent on the program.[11] Particularly during the economic downturn in the 1970s, some suburbanites opposed spending their tax dollars on integration. When Newton faced declining enrollments in the 1970s, a contingent advocated for closing a school and

[6] "Guidelines for METCO Participants," May 1966, Box 1, File 34 METCO Archives.
Bureau of Equal Educational Opportunity, "METCO Handbook," Massachusetts Department of Education, 1976, Box 1, File 39, METCO Archives

[7] "Guidelines for Metropolitan Boston Communities Interested In Participating in METCO," December 1970, Box 1, File 35, METCO Archives.

[8] "Supports School Board," Lexington Minute-Man, January 27, 1966.

[9] Bureau of Equal Educational Opportunity, "METCO Handbook," Massachusetts Department of Education, 1976, Box 1, File 39, METCO Archives.

[10] Apfelbaum, Katherine and Ardon, Ken, "Expanding METCO and Closing Achievement Gaps." Pioneer Institute White Paper No. 129. March 2015.
Joseph, Stephanie. "After George Floyd, More Suburbs Express Interest in Joining METCO." Learning Curve, WGBH. October 6, 2021. https://www.wgbh.org/news/education/2021/10/06/after-george-floyd-more-suburbs-express-interest-in-joining-metco Accessed February 21, 2022.

[11] Chanoux, Laura, "From the City to the Suburbs: School Integration and Reactions to Boston's METCO Program. Senior Honors Thesis. March 30, 2011.

consolidating enrollment to save money instead of accepting METCO students into empty seats.[12]

Connected to the cost concerns, opponents argued that METCO would increase class sizes and reduce the amount of individualized attention their students would receive from teachers. They also worried that urban students would increase the behavioral disruptions in the classroom and lower the average test scores of the school. In turn, this potential change in school quality could have a negative impact on suburban students' outcomes, perceived school quality, and in turn, housing values.[13]

Others argued that the program was discriminatory because it focused on urban students of color instead of including all urban students.[14] Some contended that the program burdened students with the long bus rides, but ignored the fact that their families willingly signed up for the program knowing the length transit time.[15] Public discourse rarely discussed whether the housing discrimination and the resulting residential segregation should be addressed, which would in turn reduce school segregation without the need for busing.[16]

Funding

Suburban districts that participate in METCO receive annual funding for the program through two sources: state education funding for public schools and a METCO-specific grant. In Massachusetts, state education funding comes from a complex formula that calculates the minimum amount per pupil that should be spent in each district. The minimum spending per pupil is based on the districts' prior year student enrollment, number of English Language Learners, and number of low-income students. Once METCO participants are enrolled for a year, they are counted in this calculation and therefore result in additional per-pupil funding from the state each year.[17]

The formula also determines how much of per pupil spending the state will pay for and how much will come from local district funding. The state pays a larger share for less affluent districts. Therefore wealthier districts receive a smaller share of their education funding from the state. The exact amount of state funding that districts receive varies by district and by year. In the 2022 fiscal year, most METCO districts received around $2,000 per pupil through state aid. Six districts received between $2,500 and $7,000 per pupil (Ardon and Hatch, 2022).

[12]A. Landsman, "Questions for C.E.E.N.," Newton Graphic, November 14, 1974, Box 44, File 16, METCO Archives.
Michael R. LeConti, "Open letter to School Committeeman ward 3 Gerald Byrne," Newton Villager, November 28, 1974, Box 44, File 15, METCO Archives.

[13]Setren (2010) finds no evidence of declines in housing prices when districts begin or expand their METCO enrollments.

[14]"House kills two bills on METCO program," Boston Globe, May 1, 1975, METCO Archives.
John W. Cole, "About Our Schools," Lexington Minute-Man, August 18, 1966, Box 53, File 30, METCO Archives.

[15]Chanoux, Laura, "From the City to the Suburbs: School Integration and Reactions to Boston's METCO Program. Senior Honors Thesis. March 30, 2011.
"Scatteration' won't help," Boston Herald, January 1966, Box 44, File 26, METCO Archives.

[16]Chanoux, Laura, "From the City to the Suburbs: School Integration and Reactions to Boston's METCO Program. Senior Honors Thesis. March 30, 2011

[17]See M.G.L. ch. 70 (2023).

The state also provides districts funding through a METCO-specific grant. This is in addition to the per pupil funding discussed above. The Massachusetts Legislature decides how much funding to appropriate to the METCO grant annually and the amount fluctuates over time. The grant amount is a function of the district's grant in the prior year and prior METCO enrollment. Prior METCO enrollment is either average enrollment in the past three years or the prior year's enrollment, whichever is larger. The grant does not go below what the district received last year and sometimes the floor was the prior amount plus $40 per pupil. Therefore schools with declines in enrollment received more per pupil. In the 2021 fiscal year, the grant per pupil ranged from $6,822 to $14,407 per pupil and the average was $7,226 per pupil (Ardon and Hatch, 2022).

The total amount of state funding that suburban districts receive for METCO, including the grant, ranged from $8,00 to $16,000 per METCO participant in the 2020 fiscal year. The median district receive $8,773 for each METCO student (Ardon and Hatch, 2022). Because state aid is not itemized into funding for METCO and suburban resident students, it is not transparent to districts how much funding they receive from the METCO program beyond the grant (Apfelbaum and Ardon, 2015).

Districts often pay the marginal cost of a student to participate in the METCO program because METCO students typically fill empty classroom seats instead of causing new classrooms to be formed. This dates back to the original design of METCO's predecessor, Operation Exodus, which was based on Boston's open enrollment policy. When suburban districts first opted into the program, they followed a similar model of accepting students for whom they had space in existing classrooms. Under this model, it is possible that districts do not incur costs of hiring new classroom teachers due to the METCO program. In addition to the marginal costs for students such as instructional materials, suburban districts need to cover the cost of transportation, special education services, and the METCO program director salary. The districts also pay for any support services they provide, such as after school transportation, programing, and additional support staff like counselors and social workers.

2 Data

To study the impact of exposure to METCO peers on suburban students, this paper uses state administrative education records from the 2001-2002 through 2019-2020 school years. These data include detailed information for all students enrolled in Massachusetts public schools such as which schools they attend, whether they participate in METCO, demographics (including race, ethnicity, gender, low-income status), special education status, English Language Learner status, standardized test scores, grade progression, attendance, suspensions, SAT and AP test taking and scores, qualifying for the Massachusetts state merit scholarship, passing the high school exit exam, and high school graduation, and students' self-reported aspirations for after high school. In addition, the data include postsecondary enrollment and degree completion for 2003 to present from the National Student Clearinghouse (NSC).

Unique student identifiers enable the linkage of students to their courses and teachers. This provides data on which courses students take, their classroom peers, and their teachers' characteristics, including race/ethnicity, years of experience, and certification. It also enables the calculation of average core class size.[18] Personnel data also include counts of METCO-related staff, guidance counselors, and other specialists.

Additional baseline demographics come from the Massachusetts Department of Public Health's birth records data. They contain birthweight, parental educational attainment, parent's marital status at birth, and whether Medicaid was used to pay for the prenatal or birth medical costs. These fields are available only for those born in Massachusetts and are used for descriptive analysis.

3 Descriptive Statistics

Resident students of METCO suburban schools are 85 percent white and score almost a half standard deviation above the average Math and English test scores in the state (see Table 1). Only six percent qualify for free and reduced price lunch. Their urban peers from the METCO program have markedly different demographics. Seventy-seven percent are Black and 17 percent are Latinx. While attending the suburban districts, METCO students score over 0.6 standard deviations lower than their suburban peers on standardized tests with an average of 0.35 standard deviations and 0.22 standard deviations below the state mean in Math and English respectively. Despite these lower test scores, METCO students score substantially higher than their peers in Boston Public Schools and Springfield Public Schools. Setren (2024) finds that while some of these differences stem from positive selection into applying for the METCO program, participating in METCO leads to large academic gains for applicants.[19]

Having peers with substantially lower test scores could negatively affect the suburban students if it changes the level of material taught in class, the amount of individual teacher attention they receive, and reduces the amount students learn from their peers. Classroom behavior is another mechanism for negative spillovers. Table 1 shows that METCO participants have almost three times higher suspension rates than their suburban peers and marginally lower attendance rates. If increased class absences or behavioral issues slow down instruction, then this could negatively affect suburban students. This paper investigates each of

[18] To determine the average class size in each school and grade, I get the class sizes for all core subject classes in the data. Then I define an individual student's class size as the largest class of any of their core classes. Finally, I calculate the average student class size for each school, grade, and year combination. Estimates are similar if instead average class size is the average size of all core classes in a school, grade, and year combination.

[19] For a detailed description of how urban students apply, get accepted, and enroll in the METCO program, see Setren (2024).

these channels to see whether there are negative spillovers on academic performance and behavioral outcomes.

On average, more advantaged school districts opted into the METCO program. Table 1 shows that districts that did not join the METCO program have higher rates of economically disadvantaged students (24 percent versus 6 percent) and lower test scores (about 0.02 standard deviation above the state mean). Table 2 illustrates the positive selection using the following OLS regression:

$$y_{gjt,i} = \beta a_{gjt,i} + \zeta_g + \gamma_t + \epsilon_{gjt,i} \tag{1}$$

Where $y_{gjt,i}$ is the outcome for student i in grade g, year t, and school j. Grade fixed effects (ζ_g) and year fixed effects (γ_t) are included. The variable a_{gjt} measures exposure to METCO peers in grade g, school j, and year t where

$$a_{gjt} = 20 * \sum_{g' \leq g} Proportion.METCO_{g'jt} \tag{2}$$

This formula for METCO exposure gives equal weight to the presence of METCO students in all grades — from grade 1 until outcome grade g. I multiply the weighted ratio by the average first grade class size in METCO districts: 20 students. That makes β, the coefficient on a_{gjt} in Equation 1, the relationship between an additional METCO student in the classroom and suburban resident student outcomes.

Columns 2 and 3 of Table 2 show that the districts that participate in METCO are positively selected because controlling for individual districts lessens the relationship between METCO presence and suburban resident outcomes. Once we restrict the sample to the suburban districts that opted into the program, we see that the districts that choose to accept relatively more METCO students are also positively selected on test scores: there is a muted relationship between proportion of METCO students in the school and suburban student test scores when we add district fixed effects (Table 2 Columns 5 and 6). The district fixed effects model shows a very small positive relationship between the proportion of METCO students and the suspension rate. This relationship could stem from METCO students increasing suspensions among resident students, higher rates of suspensions among districts that choose to accept more METCO students, or a combination of both.

After controlling for individual districts, we see that having roughly one additional METCO student per classroom is linked with scoring 0.015 standard deviations higher on Math which is much smaller than the 0.096 standard deviation difference without district controls (Table 2 Columns 5 and 6). Controlling further for lagged school demographics, it appears that higher performing schools accept more METCO students since the estimates are further reduced (Table 2 Column 7).[20] The positive relationship between

[20] The lagged school demographics results are similar to a model with school fixed effects.

METCO and suspension rates also appears for the lagged school trait model. Overall these results suggest the presence of positive selection not just into which districts participate in METCO, but which districts and schools participate relatively more in the program.

After adding controls for individual student baseline characteristics (including gender, race, ethnicity, low-income status, special education and level of services, English Language Learner and level of proficiency, and immigrant status), there is still a positive relationship between the proportion of METCO students in a school and suburban students' test scores. Having one additional METCO student on average in a class is correlated with 0.004 and 0.006 standard deviations higher on Math and English, a 0.2 percent higher attendance rate and a 0.1 percent increase in the likelihood of suspension. For suburban Black and Latinx students, the correlations are similar for English, attendance rate, and suspensions and appear slightly larger for Math (see Appendix Table 2).

The positive relationships between METCO and test scores and attendance could mean that the types of schools that enroll more METCO students also enroll suburban students that are pre-disposed to be high performing in ways that are not observable in the data. It could also reflect that the types of schools that enroll more METCO students also generate stronger test performance relative to other schools. It is also possible that METCO students generate a small positive academic effect and attendance effect on their suburban peers, particularly for students of color. Alternatively, these correlations could mask negative spillovers. The small positive relationship between suspensions and METCO students could suggest some negative spillovers.

4 Methodology

To address the positive selection of districts into accepting METCO students, this paper uses two quasi-experimental methodologies on two types of variations in METCO enrollment to identify the causal impact of METCO peers on suburban students' outcomes. The difference in differences design uses the starting, stopping, or pausing of new METCO enrollment within schools. The 2SLS method uses the fact that schools can accept more METCO students when there are more empty seats in classrooms. With an instrumental variable for predicted space in classrooms, I estimate the impact of having an additional METCO student in the classroom. The difference in difference approach estimates the effect of having any METCO peers in the grade cohort, while the 2SLS method estimates the average effect of an additional METCO peer.

For both methodologies to be valid, the fluctuations in METCO enrollment need to be unrelated to anticipated outcomes. That means in the Spring prior to enrolling in Kindergarten or first grade, suburban parents do not influence how many METCO students their child's future school enrolls. We might also be concerned about parents changing their child's school based on METCO enrollment, but whether and how many METCO students will enroll in a given year is not transparent to parents. Since the decision is made in the Spring of the prior year and timing varies by school district, parents of incoming first graders are unlikely to have influence or know exactly when the decision is made. In addition, school and district leaders do not select the number of METCO students in a cohort based on the expected outcomes of specific cohorts of students. This seems unlikely because the school does not have baseline academic outcomes for the incoming students. For the METCO enrollment to be related to potential outcomes, district leaders would need to decide whether a cohort gets METCO or not based on the potential future test scores, behavior, and other outcomes of the incoming first graders. The district leaders decide on incoming METCO enrollment the Spring of the prior year, when they do not fully know which students will enroll. They do not have any prior information about the academic preparation or behavior of the incoming class to base their decision off of.

5 Difference-in-differences & event study analysis

5.1 Methodology

I use fluctuations in METCO enrollment across cohorts within individual schools to estimate the causal impact of having METCO students as peers in a difference-in-differences analysis. Most schools in METCO receiving districts enroll METCO students at some point in time (see Panel A of Figure 1).[21] Within the schools that accept METCO students, only 68.3 percent of the cohorts have at least one METCO student. Panel B of Figure 1 shows that only three districts have METCO students in all of their cohorts. Most districts have METCO students in 35 to 85 percent of their cohorts.[22] The cohorts without METCO students reflect either that a school pauses new METCO enrollment for one or two years or they stop new METCO enrollment for several years or all together.

To utilize variation due to schools starting their METCO enrollment, I compare suburban student outcomes before and after schools start accepting METCO students. I define treatment schools as those that have switched from not enrolling METCO students in Kindergarten or first grade to enrolling METCO stu-

[21] The exceptions include the elementary and middle schools in Foxborough, Natick, and Westwood which start accepting METCO students in middle or high school. Braintree, Lincoln, and Hampden-Wilbraham have at least one elementary school each that never accepts METCO students, but they have at least one other elementary school that does.

[22] I calculate METCO representation in cohorts by using first grade enrollment numbers.

dents in subsequent cohorts.[23] By using this within school comparison, I control for school-specific attributes that do not change over the sample period. I include up to four years before the school enrolled new METCO Kindergarten or first graders and up to four years after in the analysis.

To adjust for time-varying trends, I match each treated school with comparison schools. The set of comparison schools for each treatment school includes any school in a METCO-receiving suburban district that did not accept new METCO Kindergarten or first grade students for the treatment school's analysis sample years (up to five years before and four years after the METCO enrollment policy change). This approach compares the effect of having METCO peers in the cohort to not having any METCO peers. Since it averages across all of the treated cohorts, this approach estimates the effect of the average METCO cohort size (which is 3.12 students) compared to having no METCO students in the cohort.

The sample includes all of the suburban resident students in the treatment and comparison schools and excludes METCO participants. Cohorts that have METCO students in some, but not all of their elementary school years, are excluded from the analysis. I assign students to cohorts based on what school they attended in first grade.

This process creates sets of treatment schools that started or reinstated their METCO enrollment and comparison schools that steadily had no METCO enrollment in the corresponding grade-levels and years. Each of these could be run as their own individual difference-in-differences analysis. I append the individual student-level data from each of these treated and comparison schools sets to create a stacked dataset of all treatment and comparison groups. Since schools can appear as both a treatment school and a comparison school to multiple treatment schools, I cluster standard errors on each treatment-comparison set and include indicator variables for each treatment-comparison set. See Cengiz et al. (2019) and Deshpande and Li (2019) for other examples of stacked cohort differences-in-differences analysis. This stacked difference in differences methodology corrects for the weighting problems created by staggered treatment timing discussed in Goodman-Bacon (2021). This methodology has similar properties as Sun and Abraham (2020) and Callaway and Sant'Anna (2021), the other solutions to the weighting issues of staggered treatment timing (Baker, Larcker and Wang, 2022). Stacked difference in differences is my preferred specification due to its simplicity in explaining the methodology to school leaders and policymakers.

The following estimating equation identifies the causal impact of having METCO students in a grade cohort on resident suburban student outcomes:

[23] Districts start METCO enrollment in Kindergarten or first grade. Once a METCO student is enrolled in a suburban district, they can remain enrolled until they graduate high school. There are no cases of cohorts having METCO students in Kindergarten but not first grade.

$$y_i = \beta_0 + \beta_1 TreatmentSchool_{ic} + \beta_2 Post_{ic} + \beta_3 TreatedCohort_{ic} + \delta_j + X_i + \alpha_{it} + \epsilon_{itc} \quad (3)$$

where y_i is the outcome of interest for student i. $TreatmentSchool_{ic}$ equals 1 if the student attended the school in the first grade that undergoes METCO enrollment change c by switching to accepting METCO students. $Post_{ic}$ equals 1 if individual i enrolled in first grade after METCO enrollment change c.[24] $TreatedCohort_{ic}$ equals 1 if the student attended the treated school after policy change c, and therefore had METCO peers in their grade cohort. I include school fixed effects δ_j and individual baseline covariates X_i, including gender, race, immigrant status.[25] I control for when students entered first grade with year fixed effects α_{it}. Since schools and their enrolled students can be in multiple comparison groups, I cluster standard errors at the policy change subsample.

I repeat an analogous exercise for schools that change their enrollment policy from accepting METCO Kindergarten or first graders, to stopping or pausing accepting new METCO Kindergarten or first graders. When new enrollment is stopped or paused, older METCO students can remain in the district until they graduate. Similar to the prior analysis, I include up to five years before the school enrolled METCO students and up to four years after. Suburban schools in METCO-receiving districts that had METCO enrollment in all of the cohorts during the treatment school's analysis sample years serve as the comparison schools. The estimating equation is identical except that policy change refers to ending new METCO enrollment.

In addition to checking that the treatment and comparison groups have similar baseline characteristics, I check for common pre-trends in the outcomes of interest between the treated and comparison groups. I run the following event study version of Equation 3 estimation:

$$y_i = \sum \pi_t (TreatmentSchool_{ic} * D_{ic}^t) + \sum_t D_{ic}^t + \delta_j + X_i + \alpha_{it} + \epsilon_{itc} \quad (4)$$

where the cohort dummies, D_{ic}^t control for the year t relative to the policy change c that student i enrolls in the first grade. The cohort that attends first grade in the year their school starts or resumes accepting

[24]If students repeat the first grade, I use the first attempt. If students switch schools during first grade, I use whichever school is in a suburban METCO district. If there are multiple schools in METCO suburban districts, then I assign the student to the school where they attended the most days in first grade.

[25]I exclude controls that change over time such as special education status, free and reduced price lunch status, and English Language Learner status.

METCO students have $t = 1$. The following year's first grade cohort has $t = 2$ and so on. The cohort right before the school accepts METCO students have $t = 0$ and two years before have $t = -1$. Figures 2 and 3 plot π_t for each year relative to the enrollment change. These event study figures show comparable pre-trends across treatment and comparison groups for grades 3 through 5 test scores, suspension rates, and attendance rates.

5.2 Threats to validity

Appendix Table 3 shows that both difference-in-differences analysis samples have broad coverage of districts, schools, and cohorts and similar baseline characteristics to suburban students overall (see Columns 2 and 3). In addition, over 73 percent of school districts serve in at least one treatment group.

Students in the difference-in-differences analysis samples have similar baseline characteristics as the full sample. There is no measure of baseline academic ability, but the pre-period third grade test scores of the difference-in-differences analysis samples are comparable to the full sample. Math and English test scores are about half a standard deviation above the state mean for the full sample and the analysis sample where ending the METCO program is the treatment. The analysis sample where starting METCO enrollment is the treatment has slightly lower pre-period test scores of 0.4 standard deviations above the state mean.

Appendix Table 4 shows that the demographics of the resident students in treated schools remained stable before and after the school changed whether they accepted new METCO students. Columns 2 and 5 show the coefficients from regressions of baseline characteristics on being in a cohort following a METCO enrollment policy change. The models control for school fixed effects and the sample is the set of schools that experience policy changes. These results show that the cohorts before and after METCO enrollment policy changes have minimal demographic differences. While there are statistically significant differences for Black and immigrant, all differences are less than 0.5 percentage points, so the pre and post METCO program suburban resident cohorts within schools have similar baseline characteristics.

Next, I compare the traits of all treated and comparison cohorts. Columns 3 and 6 of Appendix Table 4 show the regression of baseline characteristics on being in a treated cohort, controlling for the policy change indicators and year. The results show that suburban residents of treated and comparison schools have similar demographic characteristics, with differences all less than one percentage points. In addition, there was no differential pre-trend in baseline characteristics (see Appendix Figures 2 and 3).

5.3 Results

The intensity of the treatment varies across the treatment groups. On average, 4.1 percent of the treated grade cohorts are urban METCO students. The representation of METCO in treated school by grade cohorts ranges from 0.98 percent to 19.05 percent. This amounts to between 0.2 and 3.8 students in a 20 person class. Districts vary in the size of their METCO programs. Panel C of Figure 1 shows that METCO representation in treated cohorts ranges from 1.36 percent of the students in the district with the smallest METCO program to 8.86 percent in the largest METCO district. At the cohort level, the treatment ranges from 1 to 18 METCO students in the cohort compared to zero in the comparison group. The average treatment is 3.12 METCO students in the cohort.

Table 3 shows the results from the difference in differences regressions. Being in a cohort with METCO students has no significant effect on Math or English test scores. We can rule out Math effect sizes larger than plus or minus 0.04 standard deviations. For English, we can rule out effect sizes larger than plus or minus 0.05 standard deviations. This holds true when the treatment is a school starting METCO enrollment or ending METCO enrollment. To contextualize the magnitude of the effect sizes I can rule out, an increase of about 10 percentage point in Black enrollment share decreases test scores by 0.04 to 0.1 standard deviations in the literature (Cook, 2024; Hanushek, Kain and Rivkin, 2009; Billings, Deming and Rockoff, 2014; Hoxby, 2000).

There are also no significant effects of having METCO peers on suburban students' attendance rates: we can rule out effect sizes larger than plus or minus 2 percentage points. The attendance estimates are more precise and smaller for grade three and we can rule out effects larger than plus or minus 0.2 percentage points. The confidence intervals for suspension rate show that effects are between plus and minus 0.3 percentage points. In total, these confidence intervals allow us to rule out a substantial effect on elementary school test scores, attendance, and suspension rates as a result of having METCO students in a cohort.

The event study graphs in Figure 2 overall show null effects of having METCO peers in each of the four cohorts who enter first grade after their school starts accepting METCO students. The third and fourth grade event study graphs show no evidence of differential Math scores, English scores, attendance rates, or suspension rates for cohorts that were exposed to METCO. The fifth grade event study shows point estimates of -0.05 standard deviations in Math which are marginal significant. The other fifth grade outcomes show no statistically significant impact of cohorts having METCO exposure.

Figure 3 shows no change in student outcomes in the four cohorts that enter after their school stops accepting METCO students. This holds across grades three through five and each of the outcome variables. Estimates for the fifth and sixth year following the enrollment policy are similarly not statistically significantly different from zero, but are noisier due to a smaller sample size.

Appendix Table 5 shows that results are robust to more narrow comparison groups. Columns 2 and 6 show requiring that comparison schools are within the same district as the treated school yields similar results. Columns 3 and 7 show similar results when synthetic control weights are used for the comparison groups. The synthetic controls were fitted on the baseline characteristics of the treated cohorts to the year prior to the enrollment policy change (results are robust to assigning the synthetic control weights based on the two years prior to the policy change). Columns 4 and 8 show similar results when comparison groups come from schools that are eventually at some point in time. These specifications show mild positive effects for some outcomes.[26] Each of these alternative comparison group strategies yield similar results, though less precise point estimates due to the smaller sample size.

6 Class Size Instrumental Variable Analysis

6.1 Methodology

Teacher union contracts or district rules cap class sizes in each of the districts that accept METCO students. As a result, districts have a financial incentive to fill empty classroom seats with METCO students, but to not exceed the class size cap. Schools can lower per pupil costs by filling empty classroom seats, but exceeding the class size cap would require hiring a new teacher. I use the fact that classroom space constraints play a role in determining the number of METCO students a school accepts each year in a 2SLS method similar to Angrist and Lang (2004)'s adaptation of Maimonides' rule.

District leaders look at their expected enrollment for the following school year in the Spring, consider class size constraints, and notify METCO of how many students they want to enroll from the city for each grade-level. To instrument for the number of METCO students in a grade, I estimate the predicted class size in first grade using Maimonides' rule:

$$r_{jt} = \frac{e_{jt}}{int(\frac{e_{jt}}{maxclassj} + 1)} \tag{5}$$

where e_{jt} reflects resident, non-METCO enrollment in first grade in school j, and year t. The first grade class size limit $maxclass_j$ comes from whichever of the following measures is the lowest: the maximum class size in the district across all years, the district's teacher union contract class size maximum, or the class size limit set by district policy. The predicted first grade class size for school j and year t is represented by r_{jt}.

[26] When the treatment is starting or resuming METCO enrollment, 53.09% of the comparison schools in the main specification are treated at another point in time. When the treatment is ending METCO enrollment, 78.95% of the comparison schools in the main specification are eventually treated.

Under this formula, as schools approach enrollment numbers that are multiples of the class size cap, they add another classroom and predicted average class size shrinks.[27] Panel A of Figure 4 shows this pattern for the first grade. For illustrative purposes I set the class size cap in to 25 students, which is the largest class size cap among the METCO districts.[28] As suburban resident enrollment approaches multiples of 25, the average predicted class size sharply drops as a new classroom is added (see the dashed line in Panel A of Figure 4).

Panel A of Figure 4 shows that the Maimonides rule with a class size cutoff of 25 students closely predicts average class size for first grade enrollment of under 100 students. The prediction becomes less accurate for cohorts larger than 100 students which have more classrooms. Because of this, the model described below will control for the predicted number of classes in each school's cohort.

Another way to visualize the quality of the prediction is to plot actual average non-METCO class size against average predicted non-METCO class size, shown in Panel B of Figure 4. Since all of the data in the graph is at the school by year level for first grade, I can use the district-specific class size caps in the predicted class size instead of the generic class size cap of 25 students. Like in Panel B of Figure 4, the predicted class size is close to the actual class size in the majority of cases, with data clustering along the 45 degree line.

In Figure 5, we see that as the average class size (excluding METCO students) increases, the average number of METCO students per class falls from about 0.75 students for class sizes of 17 to close to zero students a class size of 25 students. This relationship between available classroom space and METCO students along with the strong relationship between predicted and actual class size support the premise for the instrumental variable that as the predicted class size of a school by grade cohort increases, the number of accepted METCO students decline. Therefore, I use predicted class size as an instrument for school by year METCO enrollment in first grade. To allow for a non-linear relationship between predicted class size and METCO enrollment, I also use indicator variables for integer values of predicted class size as instruments.

The second stage regression estimates the impact of having METCO peers in your cohort on individual outcomes as follows:

$$y_{gjt,i} = \tau a_{gjt} + \alpha n_{gjt} + \kappa_2 c_{gjt} + \delta_{a2} int.r_{gjt} + \delta_{b2} linear.r_{gjt} + \beta_{\alpha 2j} + \gamma_{2t} + X_i' \Gamma_2 + \epsilon_{gjt,i} \tag{6}$$

[27] Angrist and Lang (2004) also use predicted class size and integer indicator variables of predicted class size as instrumental variables in their analysis of the Brookline district's METCO program from 1994 to 2000. The also have a different specification that uses $z_{gjt} = min[max(23 - r_{gjt}, 0), 1]$ as an instrumental variable. This instrument predicts that there will be on average one METCO student per classroom if their grade's average predicted class size is 23 students or less, and none if the predicted class size is larger. This pattern is not visible in this paper's data, or Brookline in the years 2001 - 2020, the years of this study. As such, I do not use this instrument in the analysis.

[28] Except for Scituate which targets a range of 15 - 30 students per class instead of a cap.

where $y_{gjt,i}$ represents student i's outcome in grade g, school j, year t. The variable a_{gjt} measures exposure to METCO peers in grade g, school j, and year t as described in Equation 2. That makes τ, the coefficient on a_{gjt}, the impact of having one additional METCO student in the classroom on suburban resident student outcomes

I control for average non-METCO class size, n_{gjt}, in a given grade, school, and year. I add indicator variables for binned grade-level enrollment $c_{gjt} = int(e_{gjt}/25+1)$, which proxy for the number of classrooms in the grade level. This allows the predictive power of the instrument to vary non-linearly and adjusts for the fact that the instrument has lower predictive power for larger cohorts (shown in Figure 4). I also control for the instrumental variables: integers of the predicted class size variable $int.r_{gjt}$ and predicted class size $linear.r_{gjt}$. I include school and year fixed effects, $\beta_{\alpha 2j}$ and γ_{2t} as well as individual baseline covariates $X_i^{'}$ which include race, ethnicity, and sex indicators, free and reduced price lunch status, special education status, and English language learner status.

The first stage equation is:

$$a_{gjt} = \beta_{1j} + \gamma_{1t} + \delta_{a1} int.r_{gjt} + \delta_{b1} linear.r_{gjt} + \kappa_1 c_{gjt} + X_i' \Gamma_1 + \nu_{gjt,i} \tag{7}$$

$$n_{gjt} = \beta_{1j} + \gamma_{1t} + \delta_{a1} int.r_{gjt} + \delta_{b1} linear.r_{gjt} + \kappa_1 c_{gjt} + X_i' \Gamma_1 + \nu_{gjt,i} \tag{8}$$

Both the proportion of METCO students, a_{gjt}, and the average number of non-METCO students per class, n_{gjt}, are endogenous variables. Higher performing schools may be pre-disposed to have smaller class sizes and accept more METCO students. Since all of the instruments are non-linear functions of cohort enrollment e_{gjt}, estimates that treat average non-METCO class size as endogenous may be imprecise. However, we can reject the null hypothesis that the model is under-identified with the linear and non-linear predicted class size as instruments (p-value of the Anderson canonical correlation LM statistic is 0.000 for all outcome variables). Therefore, the linear and non-linear predicted class size instruments satisfy the rank condition.

I estimated the first stages separately for each individual school district with a balanced panel sample that had grade 3 through 5 outcomes. I include the 19 districts with elementary school entry, enough predicted class size variation to have an identified first stage, and strong first stage F-statistics in the instrumental variable analysis.[29]

[29] Of the 38 suburban districts that accept METCO students during the study period, two districts do not have any elementary schools and four do not have elementary school entry points. Seven districts do not have enough predicted class size variation to have an identified first stage. Five additional districts did not have enough predicted class size variation to have an identified first stage when I restrict the sample to a balanced panel with 3rd through 5th grade outcomes. Lastly, two districts were dropped for having weak first stages with F-statistics of 16 and 64. These were clear outliers relative to the other districts which had a median F-statistic of 224 and the 25th percentile was 178. Combined, those limitations yield 19 districts for the instrumental variable analysis. Results are robust to including the two districts with weaker first stages in the analysis.

6.2 Threats to validity

For the predicted class size instruments to meet the monotonicity assumption, we need increases in predicted class size to have either a neutral or negative effect on the average number of METCO students in class for each individual school and first grade cohort. This seems plausible because it seems unlikely that schools accept more METCO students when faced with less classroom space.

The independence assumption requires that fluctuations in the first grade cohort size (the variable input in the predicted class size formula) has a "as good as random" relationship with the relevant observable and unobservable variables. To test the relationship with observables, I regress each of the baseline covariates on predicted class size and coefficients are all close to zero (see Appendix Table 6). The coefficients are under 0.3 percentage points for the indicator baseline variables and less than 2.5 grams for infant birthweight compared to the mean of 3,390 grams. Column 6 shows that most coefficients are not statistically significant, but those that are are precise small coefficients of 0.0 to 0.3 percentage points or under 2.5 grams. This assumption also requires that unobservables like racial attitudes of the child or their family, home environment, and motivation are not correlated with cohort size.

The exclusion restriction requires that predicted class size only affects student outcomes through differences in METCO exposure. However, predicted class size and the treatment of increased METCO exposure also effects actual class size which in turn affects student outcomes (see Hanushek (2003) for review of evidence). We can consider the increased class size as a result of METCO a part of the METCO exposure treatment. Additionally, this problem is bounded by the magnitude of the class size changes observed in the data: predicted class sizes range from 17 to 25 students.

While the state administrative data spans from 2001 - 2020, it only has class size data for 2012 - 2020. This presents a challenge in estimating Equation 6 for the earlier years without non-METCO class size, n_{gjt}. To adjust for this, I use data from 2012 - 2020 to estimate the following equation:

$$a_{gjt} = \beta_{1j} + \gamma_{1t} + \delta_{a1} int.r_{gjt} + \delta_{b1} linear.r_{gjt} + \kappa_1 c_{gjt} + X_i' \Gamma_1 + \nu_{gjt,i} \tag{9}$$

Then I use those estimated regression coefficients to predict average non-METCO class size, $\hat{n}_{gjt}$ for 2001 - 2011. Finally I estimate the second stage regression, equation, on the full sample, but using $\hat{n}_{gjt}$ for the earlier years.[30] Results are robust to using just 2012-2020 data, but estimates are more precise with the full sample (see Appendix Table 7). Lastly, Appendix Table 3 shows that the sample of districts in the 2SLS analysis have similar average characteristics to the full sample of suburban residents.

[30] See Pacini and Windmeijer (2016) for a discussion of the two-sample 2SLS estimator and its properties.

6.3 Results

Similar to the other identification strategy, I find no evidence of a significant effect of METCO students on their suburban peers from the two stage least squares estimates (see Table 4). The 2SLS estimates are less precise than the difference in differences estimates. The first stage for proportion of students in METCO have Angrist and Pischke F-stats of over 100, signaling a valid first stage. I find that the addition of METCO student in a classroom leads to Math test score changes close to zero. The effect size is 0.02 standard deviations in third grade, 0.036 standard deviations in fourth grade, and 0.009 standard deviations in fifth grade. The estimates are similarly small for English, ranging from -0.016 standard deviations to 0.067 standard deviations. The standard errors for each grade level are around 0.03 standard deviations. I can rule out effects larger than -0.05 standard deviations on Math scores and -0.04 standard deviations in English.

For attendance, I can rule out effects larger than a 0.3 percentage point drop for grades 3 through 5, a 1.6 percentage point drop for grade 2, and a 0.1 decline for first grade. Lastly, the two stage least squares estimates can rule out effects above a 0.45 percentage point increase in the suspension rate. In all, I can rule out substantial negative peer effects from adding an additional METCO student to a classroom from the instrumental variable estimates.

7 Effects of enrollment, classroom, teacher characteristics, and longer run outcomes

7.1 School and district switching

We might be concerned that METCO enrollment could alter families' enrollment decisions. This could impact school funding and could result in biased estimates if students are more likely to leave the school or district when the school starts, stops, or pauses METCO enrollment. Table 5 shows the stacked difference in differences estimates of METCO enrollment on switching school for elementary school grades. Column 1 shows that 14 percent of suburban students in schools and cohorts without METCO switch schools to another school within their district by fifth grade. Students in cohorts with METCO are not statistically significantly more likely to switch. The effect sizes for school switching in grades one through four are less than 0.6 percentage points and the effect for grade five is less than 1.5 percentage points. I can rule out school switching of more than 0.6 percentage points in grades one and two, more than 2.4 percentage points for grades three and four, and more than 4.4 percentage points for grade five.

Similarly, students are not substantially more likely to leave the school district for another public district, charter school, private school, or out-of-state school when their cohort has METCO students. Column 3 shows that 13.7 percent of students in non-METCO cohorts leave the district by the fifth grade. Those with METCO in their cohort are 0.7 percentage points more likely leave to the district by the fifth grade. The effect sizes are smaller for the earlier grades. I can rule out effects larger than 1.5 percentage points.

Columns 5 through 8 show analogous estimates for when schools pause or stop METCO enrollment which yields similar findings. Estimates are less than 0.2 percentage points for grades one and two. Grade three and four estimates are noisy and around 1 percentage point each. Not having METCO in the grade cohort reduces the likelihood of switching schools within the district by 2.4 percentage points by fifth grade. Students are similarly likely to leave the school district when they do not have METCO enrollment in their grade cohort. I can rule out effect sizes larger than 1.2 percentage points. Columns 9 through 12 show similar null effects for adding an additional METCO student to the classroom through the instrumental variable approach. Overall, this exercise shows that having METCO students in the cohort has a small and mostly statistically insignificant effect on students leaving the school or district.[31]

7.2 Impact of METCO on class traits

Next, I explore whether cohorts with METCO students have different classroom and teacher characteristics than those without. The METCO program may attract teachers to the suburban districts if they are excited about teaching more diverse classes. Alternatively, it may deter teachers if they find teaching classes with METCO students challenging (for example, if there are changes in classroom behavior or the ability distribution). As a result, the METCO program may impact the types of teachers suburban students have.

Difference-in-difference estimates in Table 6 show that school participation in METCO has minimal effect on the teacher characteristics that a cohort experiences in grade four (results are similar for other grade levels). Teacher qualifications, years of experience, and race are largely similar when schools start or stop accepting METCO students. It appears Hispanic teachers are slightly less present in core classes when METCO enrollment pauses or stops.

To investigate whether the presence of METCO students in a cohort affects the distribution of ability levels in a class, I look at the average difference between the 90th and 10th percentiles' lagged test scores

[31]If students switch schools to another Massachusetts public school, they remain in the data.

in students' average core classes. If enrolling METCO students leads to more tracking, we would expect a narrowing in the lagged test score distribution of classes and a negative point estimate for when METCO programs start. There appears to be no significant change in tracking or the range of ability levels in core classes when schools resume or start METCO enrollment.

METCO cohorts have on average 0.75 more METCO students per 20 student class. The representation of special education and English Language Learners are similar across cohorts with and without METCO. The presence of economically disadvantaged students' increases slightly in METCO cohorts. METCO enrollment reduces class size by about 0.68 students when schools switch to accepting METCO students and increases average class size by about 0.45 students when districts stop accepting METCO students. Since districts often use METCO to fill empty seats, this suggests that districts only partially fill empty seats with METCO students. This is also demonstrated by the first stage being less than one student for one additional seat in the instrumental variable analysis. It is possible that the reduction in class size benefits students and counteracts any potential negative effects from additional METCO peers. Lastly, the suspension and attendance rates in classes, both overall and among resident students, appear similar across METCO enrollment status. Therefore, the presence of METCO students in a cohort does not significantly change the classroom or teacher characteristics beyond the demographic changes from having the METCO students and slightly decreasing class size.

Columns 7 through 9 show effect of an additional METCO student in a classroom on the teacher and classroom traits using the instrumental variables approach. An additional METCO student is linked with slightly lower average teacher experience and small shifts in teacher racial composition (a 2.6 percentage point increase in the likelihood of a white teacher, a 1.1 percentage point decrease in Black teachers, and a 3.6 percentage point decrease in Asian teachers). An additional METCO student increases the class distribution of ability levels in a class by 0.123 standard deviations in Math and 0.184 standard deviations in English. This shows that METCO does not lead to more tracking, but slightly increases the range of ability levels in the classroom. An additional METCO student increases the proportion of low-income students in the classroom by 1.7 percentage points and slightly reduces the rate of special education students in the classroom. An additional METCO student in the classroom is linked with an average class size reduction of 2.5 students. As discussed above, this is driven by schools accepting more students if they anticipate smaller class sizes, but not fully filling all empty seats.

7.3 Impact of METCO on high school and college outcomes

People who are concerned about potential negative impacts of increased racial and socioeconomic integration might be more concerned about college preparation and college going outcomes than test score impacts. The relatively high performing suburban schools have high test scores. For example, Table 7 Column 1 shows that 93.4 percent of students without METCO classmates pass the exams required to graduate high school and Table 1 showed that on average students perform over 0.4 standard deviations above the state mean. Therefore, opponents of increased diversity might worry more about advanced coursework or competitive college admissions.

Tables 7 and 8 show the two-stage least squares estimates for high school and college outcomes.[32] Column 5 in each table shows strong first stage F-statistics for the first grade predicted class size instrument even into high school. This reflects how all but one of these districts only have one high school each. METCO students that remain in the program through high school will remain in the same grade cohort as their initial first grade peers since they typically enroll in the same middle school and high school. As a result, the instrumental variables based on predicted first grade class size continue to have predictive power of METCO exposure in high school. The difference-in-differences design is less strong for later grades because students in first grade cohorts without METCO often have METCO students in their grade cohort when they enroll in middle or high school. This happens because districts often have multiple elementary schools that feed into fewer middle schools.

Table 7 shows that there are no substantive negative effects of having more METCO peers on high school outcomes. The estimates for meeting the high school testing graduation requirement are close to zero (0.7 percentage points more likely to pass but not statistically significant). The effect on qualifying for a state merit scholarship is also positive and close to zero. There are small positive and significant effects on SAT taking and scores. The positive SAT effects could be due to positive spillovers of encouragement METCO students receive to take and prepare for the SAT from the non-profit that runs METCO and the METCO Director in the school. There is a very small positive effect on the number of APs taken and passing an AP exam. The effects on dropout and high school graduation are close to zero (less than 0.2 percentage points) and there is no effect on college aspirations.

Table 8 shows that having more METCO peers does not induce students to attend a more diverse college.

[32] I show just the two-stage least squares method because cohorts from different elementary schools combine in middle and high school. As a result, the comparison cohorts from the difference-in-differences strategy often get METCO students and only a subset of the comparison groups for the main analysis work for longer term outcomes.

Students with more METCO exposure are marginally more likely to enroll in a four-year college, which is driven by private college enrollment. Students appear more likely to enroll in more competitive colleges by a small margin. For example, students are 2.3 percentage points more likely to enroll in a "very competitive" college according to Barron's rankings. There are also small positive effects on persistence and college graduation. Students with more METCO peers are 2.6 percentage points more likely to graduate four-year college (a 4.8 percent increase compared to those with no METCO peers). These positive effects on college outcomes could come from positive spillovers of METCO peers and supports. METCO students receive help with the college application process. This encouragement and support through the process could help suburban resident students if the METCO students discuss their progress in applying to college. In all, I can conclusively rule out substantial negative effects of having additional METCO peers on high school and college outcomes and there appear to be small positive effects.

8 Heterogeneity

8.1 Classmates of METCO students

Average effects could mask heterogeneity, especially if certain types of students have more exposure to METCO students in their academic core classes or if certain types of students are more affected by METCO exposure (as Sacerdote (2011) notes) . Within a cohort, some students may have more METCO peers in their classes than others. Schools often group METCO students together so that they are not the only METCO student in their class. In addition, schools may assign METCO students to classes with certain student, teacher, and curricular characteristics. If only a subset of suburban students have METCO students in their classes, then the overall analysis could mask heterogeneous effects. Classroom-level data shows that on average, 47 percent of the core classes in grade cohorts with METCO students have at least one METCO student. Of the cohorts with METCO students, only five percent of cohorts have METCO students in every core class. Forty-three percent of suburban resident students have METCO students in at least one class in a given year. Thirty-two percent of resident students with METCO students in their cohort never have a METCO student in one of their core classes in grade K through 12.[33]

Another identification strategy that accounts for how suburban and METCO students are assigned to classes would be needed to estimate differential effects by classroom exposure. Since the classroom assignment processes vary across schools and are often non-random it is difficult to properly estimate the causal impact

[33] Author's calculations.

of METCO peers at the classroom level. Instead, I can describe which students have more METCO peers in their core classes to understand who receives more exposure to METCO peers. Then, I can see if there are heterogeneous effects across these student traits. Through this exercise, I can also describe how the teacher and classroom characteristics of those exposed more to METCO vary from classrooms without METCO students.

Table 9 compares the traits of core classes with and without METCO students in schools and grade cohorts with METCO. Column 1 shows the average traits of classes without METCO, while Columns 2 and 3 display the difference between classes with and without METCO students, controlling for school, grade, and year fixed effects. Teacher qualifications appear similar among teachers who teach METCO students compared to teachers of classes without METCO students: teacher licensure rates, the state's "highly qualified" status, advanced degree rates, and average years of teaching experience are all similar. Classes with METCO are significantly more likely to be taught by Latinx or Black teachers and less likely to be taught by white or Asian teachers, though the differences are small. For example, 2.2 percent of classes with METCO students are taught by a Black teacher compared to 1.2 percent of classes without METCO students. The average value-added of teachers is 0.144 standard deviations higher in classes with METCO students than those without for Math, but teacher value-added is not statistically significantly different for English.[34]

Suburban students in classes with METCO students scored similarly to those in classes without METCO students.[35] Classes without METCO students have about a 1.8 standard deviation ability distribution (when measured as the difference between the 90th and 10th percentile of students' lagged test scores). Students in classes with METCO students experience a significantly wider skill distribution by 0.174 standard deviations wider in Math and 0.137 standard deviations wider in English. A small portion of this is due to METCO students having on average lower test scores than suburban resident students. The ability distribution of the resident students in classes with METCO students shows that on average schools place METCO students in classes with wider ability distributions among the non-METCO students. Low-income suburban residents are also over-represented in classes with METCO students.

While there is a wider ability distribution, classes with METCO kids are less likely to have classmates who receive special education or English Language Learner services. Classes with METCO students are on average 1.4 students larger than those without. Classrooms without METCO students have a 0.4 percent suspension rate. Those with METCO have a 0.1 percentage point higher suspension rate. This difference

[34]Teacher value-added is estimated on suburban resident students using a model with teacher and class-level random effects. Controls include lagged centered test scores, demographics (female, Black, Latinx, free or reduced price lunch), special education, English learner status, and year of birth. All test scores are centered to the state average in that year and grade. Value-added is only available for Math and English teachers in grades 4 - 8 and 10 (when tests are administered and there is a lagged test score). Estimates leave out data from the year they are estimating.

[35]When the number of METCO students in the classroom are taken into account, classes with more METCO students have on average resident students with lower baseline test scores.

mostly due to the relatively higher suspension rate of METCO students compared to suburban resident students since the lagged suspension rate of residents in the class is less than 0.1 percentage points different. Classrooms with METCO students have very similar, but slightly higher attendance rates by 0.5 percentage points. Lastly, core classes with METCO students have five percentage points more Black students and 1 percentage point more Latinx students (compared to two and six percent respectively in classes without METCO students). Most of the increased diversity comes from METCO itself, and not from also placing resident Black and Latinx students in classes with METCO peers.

Students with more exposure to METCO or students with certain characteristics may experience a different effect from having METCO peers. Angrist and Lang (2004) find suggestive evidence that having METCO peers may be detrimental to Black girls and Sacerdote (2011) summarizes how peer effects are not linear-in-means. Appendix Tables 9 and 10 show individual subgroup effects for both methods. The 2SLS Math and English test score estimates are mostly imprecise. Estimates for Black or Latinx girls are positive, while the effects for Black or Latinx boys are positive for English but not Math. Estimates for white or Asian boys are also positive, but the estimate for white or Asian girls is negative for Math. Low-income and non-low-income students have positive effects and the English effects are significant for low-income students.

The small number of Black and Latinx resident students in the suburban districts made it infeasible to estimate the difference in difference specification for those subgroups. Appendix Table 10 shows the subgroup estimates for the other categories. They are not statistically significant and are close to zero.

There is no strong evidence that economically disadvantaged students, Black, or Latinx students, who are more likely to be in classes with METCO students, experience significant negative effects from the program. The 2SLS analysis suggests a positive and larger effect of METCO peers for low-income students.

8.2 District-level and time heterogeneity

One strength of studying the impacts of increased diversity in this setting is the ability to see its impacts across 38 school districts and two decades. Districts have varying levels of resident diversity, demographics, academic outcomes, socioeconomic factors (see Table 1 and Appendix Table 8). Districts range from 0 to 9 percent Black, 0 to 17 percent Latinx, 0 to 45 percent Asian, and 41 to 99 percent white. Some districts have hardly any low-income students while others are almost one third low-income. Some districts' residents score close to the state average on exams, while others are 0.8 and 1 standard deviation above the state average in English and Math respectively. Educational attainment, homeownership rates, and median household income vary across the towns (see Appendix Table 8). Lastly college achievement varies across schools from

50 percent of students graduating a 4-year college to 79 percent.

Appendix Figure 4 shows the two-stage least squares estimates for individual districts that have a significant first stage (F-statistic is above 100) for third grade outcomes. There is very limited heterogeneity across school districts. Math test score effects show two districts with statistically significant negative effects and two with statistically significant positive effects. For most districts, the estimates are close to zero with a standard error of about 0.08 standard deviations. English results are similar with minimal significant variation across the districts. For attendance rate, we can rule out effects of larger than plus or minus 1 percentage points for all but two individual districts. Similarly, suspension rate effects are mostly small and close to zero for individual school districts. Effects are similar for the other elementary school grade levels. This exercise shows no substantial heterogeneities in the impact of having additional METCO peers in the classroom on test scores, attendance, or behaviors across school districts.

Next I estimate subgroup effects for different district characteristic and time periods for both identification strategies (see Appendix Tables 11 and 12). There are no substantial differences between effects across earlier versus later years in the data. Districts with above average test scores have negative 2SLS point estimates for Math and English effects and districts with below average have positive point estimates. However, these estimates are not statistically significantly different from zero and the difference in differences estimates have the reverse sign. Similarly, the estimates for districts with the most and least proportion of METCO students and most and least integration supports do not show consistent, statistically significant results across specifications and outcomes. Lastly, the six individual districts that have consistent negative (though noisy) point estimates across each of the outcomes have different key traits. Half have below average integration supports and half have above average supports. Some have a small to moderate share of students from the METCO program while others have a relatively large share. In all, there is no clear pattern of differential effects across district characteristics. We can rule out substantial negative effects of increasing students of color to by 1 to 10 percentage points of the grade cohort across a variety of districts.

9 Conclusion

Efforts to increase racial and socioeconomic diversity in schools and neighborhoods through school assignment or housing policy are often met with concern about negative peer effects. Using two different margins of increased diversity, having urban students of color in the grade cohort and having additional urban students of color in the classroom, I can rule out substantial negative peer effects on test scores, attendance, or suspension. I also find no meaningful effects on the classroom rates of suspension or classroom ability distribution that the suburban students experience. Therefore, there is no evidence that students are more likely to experience classroom disruptions or different levels of rigor as a result of the increased diversity. Lastly, I can rule out negative effects on college preparation, enrollment, persistence, and graduation outcomes. The competitiveness of the college students attend does not decline as a result of increased diversity.

There are not substantial heterogeneities across district, student characteristics, or size of the integration program. This suggests that increases ranging from one to 10 percentage point increases in urban students of color in suburban schools has no substantive effect on suburban student academic or behavioral outcomes across 38 districts. two metro-areas, and 20 years. In all, the findings suggest that concerns of negative academic and behavioral effects from up to 10 percentage point increases in diversity are not substantiated by data. However there may be effects on other important outcomes. METCO aims to broaden students' appreciation for backgrounds different from their own, build inter-racial friendships, and prepare students for diverse environments. Future work will look at the social, civic, and intergenerational impacts of increased exposure to diversity.

References

Advisory Committee on Racial Imbalance and Education. 1965. "Because It Is Right Educationally." *Massachusetts Department of Elementary and Secondary Education.*

Angrist, Joshua, and Victor Lavy. 1999. "Using Maimonides' Rule to Estimate the Effect of Class Size on Student Achievement." *Quarterly Journal of Economics*, 114(2): 533–575.

Angrist, Joshua D, and Kevin Lang. 2004. "Does School Integration Generate Peer Effects? Evidence from Boston's Metco Program." *American Economic Review*, 94(5): 1613–1634.

Apfelbaum, Katherine, and Ken Ardon. 2015. "Expanding METCO and Closing Achievement Gaps." *Pioneer Institute White Paper No. 129.*

Ardon, Ken, and Roger Hatch. 2022. "METCO Funding: Understanding Massachusetts' Voluntary School Desegregation Program." *Pioneer Institute White Paper*, , (251).

Ashenfelter, Orley C. 2006. "Evaluating the Role of Brown v. Board in School Equalization, Desegregation, and the Income of African Americans." *American Law and Economics Review*, 8: 213–248.

Baker, Andrew, David F. Larcker, and Charles C. Y. Wang. 2022. "How Much Should We Trust Staggered Difference-In-Differences Estimates?" *Journal of Financial Economics*, 144(2): 370–395.

Batson, Ruth M., and Robert C. Hayden. 1987. *A History of METCO.* The Metropolitan Council for Educational Opportunity.

Baum-Snow, Nate, and Byron Lutz. 2011. "Changes in Residential Location Patterns by Race." *American Economic Review*, 101: 3019–3046.

Belsha, Kalyn. 2023. "Feds Launch $10 Million School Desegregation Program After Stops and Starts." *Chalkbeat*.

Belsha, Kalyn, and Sarah Darville. 2020. "A New National Effort to Promote School Integration is Underway. More Than Two Dozen School Districts Want in." *Chalkbeat*.

Bergman, Peter. 2018. "The Risks and Benefits of School Integration for Participating Students: Evidence from a Randomized Desegregation Program." Institute for the Study of Labor (IZA).

Billings, Steve B., David J. Deming, and J. Rockoff. 2014. "School Segregation, Educational Attainment, and Crime: Evidence from the End of Busing in Charlotte-Mecklenburg." *Quarterly Journal of Economics*, 129: 435–476.

Callaway, Brantly, and Pedro H. C. Sant'Anna. 2021. "Difference-in-Differences with Multiple Time Periods." *Journal of Econometrics*, 225(2): 200–230.

Card, David, and Jesse Rothstein. 2007. "Racial Segregation and the Black-White Test Score Gap." *Journal of Public Economics*, 91(11-12): 2158–2184.

Cengiz, Doruk, Arindrajit Dube, Attila Linder, and Ben Zipperer. 2019. "The Effect of Minimum Wages on Low-Wage Jobs." *The Quarterly Journal of Economics*, 134(3): 1405–1454.

Clotfelter, Charles T. 1999. "Public School Segregation in Metropolitan Areas." *Land Economics*, 75(4): 487–504.

Cohen, Muriel. 1990. "After 25 Years, Lincoln Takes a Hard Look at METCO." *Boston Globe*, 21.

Cook, Jason. 2024. "Race-blind Admissions, School Segregation, and Student Outcomes." *Journal of Public Economics*, 239.

Deshpande, Manasi, and Yue Li. 2019. "Who Is Screened Out? Application Costs and the Targeting of Disability Programs." *American Economic Journal: Economic Policy*, 11(4): 213–248.

Domina, Thurston, Devon Carlson, James III Carter, Matthew Lenard, Andrew McEachin, and Rachel Perera. 2021. "The Kids on the Bus: The Academic Consequences of Diversity-Driven School Reassignments." *The Journal of Policy Analysis and Management*, 40(4): 1197–1229.

Eaton, Susan. 2001. *The Other Boston Busing Story.* New Haven, CT:Yale University Press.

Ellen, Ingrid Gould. 2000. *Sharing America's Neighborhoods: The Prospects for Stable Racial Integration.* Cambridge, MA: Harvard University Press.

Formisano, Ronald P. 1991. *Boston Against Busing: Race, Class, and Ethnicity in the 1960s and 1970s.* University of North Carolina Press.

Gamoran, A., and B. P. An. 2016. "Effects of School Segregation and School Resources in a Changing Policy Context." *Educational Evaluation and Policy Analysis*, 38(1): 43–64.

Godsil, Rachel D, Linda R Tropp, Phillip Atiba Goff, and John A Powell. 2014. "Addressing Implicit Bias, Racial Anxiety, and Stereotype Threat in Education and Health Care." *The Science of Equality*, 1.

Goldstein, Dana. 2019. "Where Civility Is a Motto, a School Integration Fight Turns Bitter." *The New York Times.*

Goodman-Bacon, Andrew. 2021. "Difference-in-Differences with Variation in Treatment Timing." *Journal of Econometrics*, 225(2): 254–277.

Government Accountability Office. 2016. "K-12 Education: Better Use of Information Could Help Agencies Identify Disparities and Address Racial Discrimination." *Report GAO-16-345*, Washington, D.C.: U.S. Government Printing Office.

Guryan, Jonathan. 2004. "Desegregation and Black Dropout Rates." *The American Economic Review*, 94(4): 919–943.

Hanushek, Eric. 2003. "The Failure of Input-based Schooling Policies." *The Economic Journal*, 113: F64–F98.

Hanushek, Eric, John Kain, and Steve Rivkin. 2009. "New Evidence About Brown v. Board of Education: The Complex Effects of School Racial Composition on Achievement." *Journal of Labor Economics*, 27(3): 349–383.

Hill, Darryl V., Rodney P. Hughes, Matthew A. Lenard, David D. Liebowitz, and Lindsay C. Page. 2023. "New schools and new classmates: The disruption and peer group effects of school reassignment." *Economics of Education Review*, 92.

Hoxby, Caroline. 2000. "Peer Effects in the Classroom: Learning from Gender and Race Variation." *NBER Working Paper Series No. 7867.*

Johnson, Rucker C. 2015. "Long-run Impacts of School Desegregation & School Quality on Adult Attainments." *NBER Working Paper Series No. 16664.*

Levy, Frank. 1971. *Northern Schools and Civil Rights; The Racial Imbalance Act of Massachusetts.* Chicago: Markham Publishing Co.

Liebowitz, David D., and Lindsay C. Page. 2014. "Does School Policy Affect Housing Choices? Evidence From the End of Desegregation in Charlotte-Mecklenburg." *American Educational Research Journal*, 51(4): 671–703.

Logan, John R., Brian J. Stults, and Reynolds Farley. 2004. "Segregation of Minorities in the Metropolis: Two Decades of Change." *Demography*, 41(1): 1–22.

Logan, John R., Deirdre Oakley, and Jacob Stowell. 2008. "School Segregation in Metropolitan Regions, 1970–2000: The Impacts of Policy Choices on Public Education." *American Journal of Sociology*, 113(6): 1611–44.

Lutz, B. 2011. "The End of Court-ordered Desegregation." *American Economic Journal: Economic Policy*, 3: 130–168.

Martin, Naomi. 2021. "Legislative Bills Take Aim at School Segregation in Massachusetts." *Boston Globe.*

McGuirk, Brendan. 2019. "A Modern METCO." *Dig Boston.*

Office of Elementary & Secondary Education. 2023. "Fostering Diverse Schools Demonstration Grants Program."

Pacini, David, and Frank Windmeijer. 2016. "Robust Inference for the Two-Sample 2SLS Estimator." *Economic Letters*, 146: 50–54.

Potter, Halley, and Michelle Burris. 2020. "Here Is What School Integration in America Looks Like Today." *The Century Foundation Early Education.*

Reardon, Sean F., John T. Yun, and Tamela Mcnulty Eitle. 2000. "The Changing Structure of School Segregation: Measurement and Evidence of Multiracial Metropolitan-Area School Segregation, 1989-1995." *Demography*, 37(3): 351–64.

Reber, Sarah J. 2005. "Court-Ordered Desegregation Successes and Failures Integrating American Schools since Brown versus Board of Education." *The Journal of Human Resources*, 40(3): 559–590.

Reber, Sarah J. 2010. "School Desegregation and Educational Attainment for Blacks." *Journal of Human Resources*, 45(4): 893–914.

Saatcioglu, A. 2010. "Disentangling School and Student-Level Effects of Desegregation and Resegregation on the Dropout Problme in Urban High Schools: Evidence from the Cleveland Municipal School District, 1977-1998." *Teachers College Record*, 112(5): 1391–1442.

Sacerdote, Bruce. 2011. "Peer Effects in Education: How Might They Work, How Big Are They and How Much Do We Know Thus Far?" In *Handbook of the Economics of Education.* Vol. 3, Chapter 4, 249–277. Elsevier.

Setren, Elizabeth M. 2024. "Busing to Opportunity? The Impacts of the METCO Voluntary School Desegregation Program on Urban Students of Color." *NBER Working Paper 32864*.

Sun, Liyang, and Sarah Abraham. 2020. "Estimating Dynamic Treatment Effects in Event Studies with Heterogeneous Treatment Effects." *Journal of Econometrics*.

Town of Brookline Board of Selectmen. 2014. "Public Hearing on Override Study Committee Report."

Tutino, Ann. 1966. "Young Bars Special Role For Negroes: Treat Them as Any Others, School Chief Declares." *Braintree Sunday News*.

Tuttle, Cody. 2019. "The Long-run Economic Effects of School Desegregation." *Working Paper*.

Vigdor, Jacob, and Jens Ludwig. 2007. "Segregation and the Black-White Test Score Gap." *NBER Working Paper 12988*.

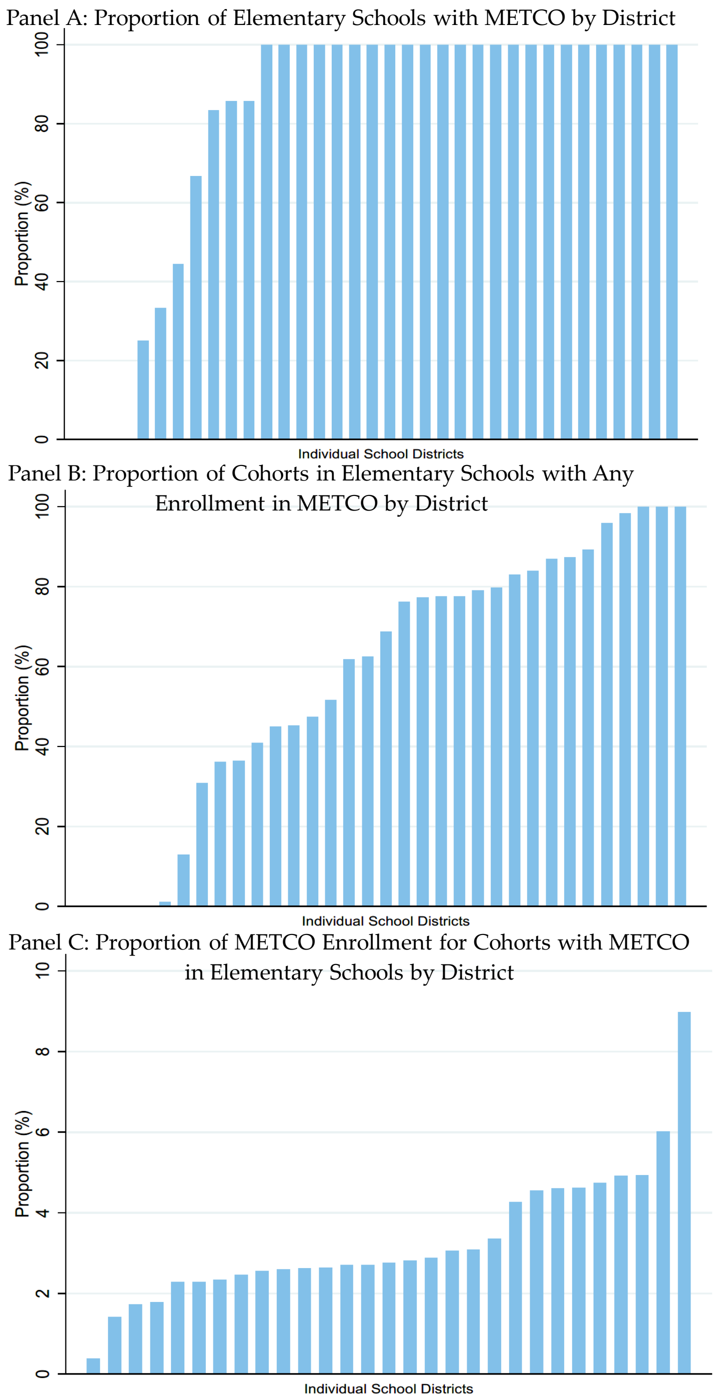
Vigue, Doreen Iudica. 1999. "Metco Students May Be Ousted from Lynnfield." *Boston Globe*, Section A: 1.

Welch, F., and A Light. 1987. "New Evidence on School Desegregation." *Technical Report, United States Commission on Civil Rights, Washington*.

Wells, Amy Stuart. 2015. "Diverse Housing, Diverse Schooling: How Policy Can Stabilize Racial Demographic Change in Cities and Suburbs." *National Education Policy Center*.

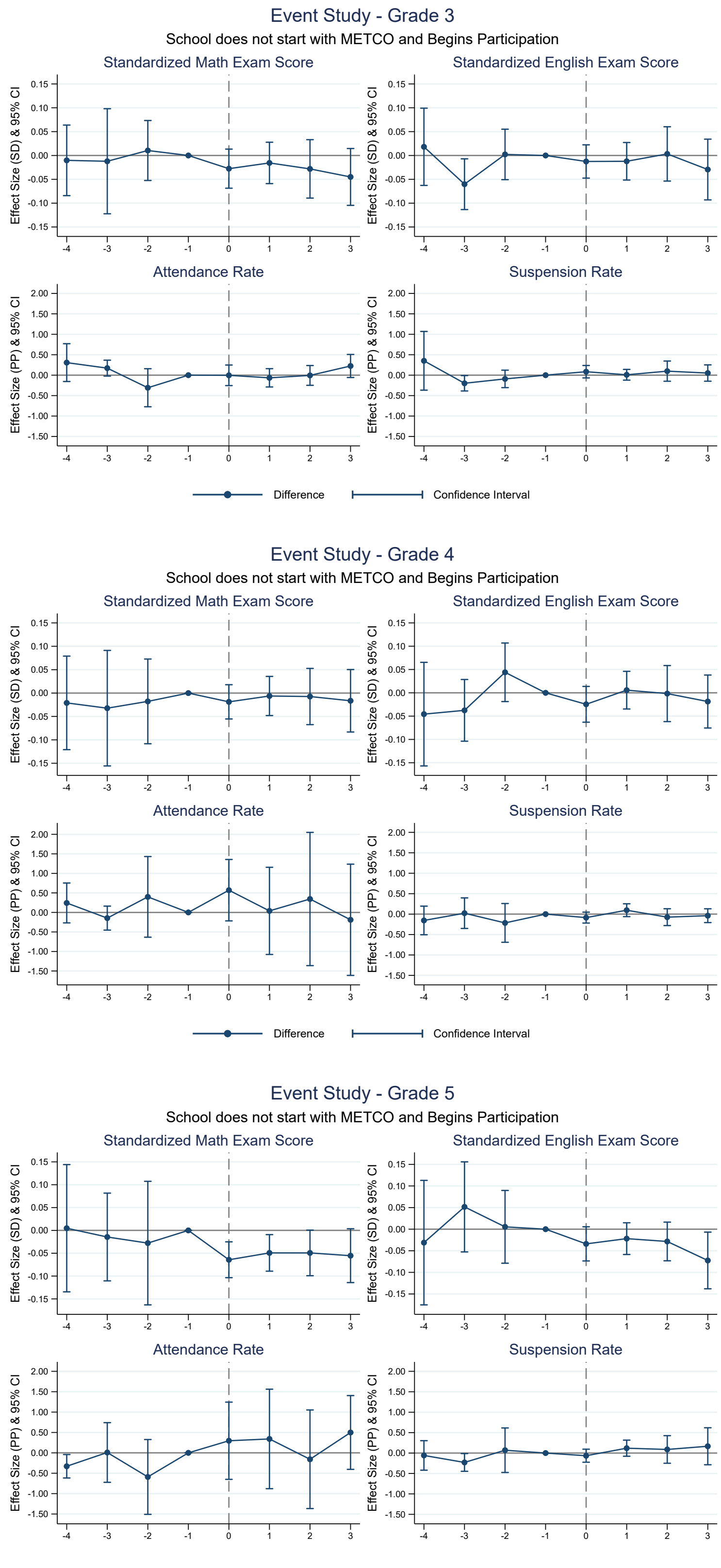
Zebrowitz, Leslie A, Benjamin White, and Kristin Wieneke. 2008. "Mere Exposure and Racial Prejudice: Exposure to Other-Race Faces Increases Liking for Strangers of that Race." *Social Cognition*, 26(3): 259–275.

Figure 1: Proportion of Elementary Schools in District that Ever Have METCO



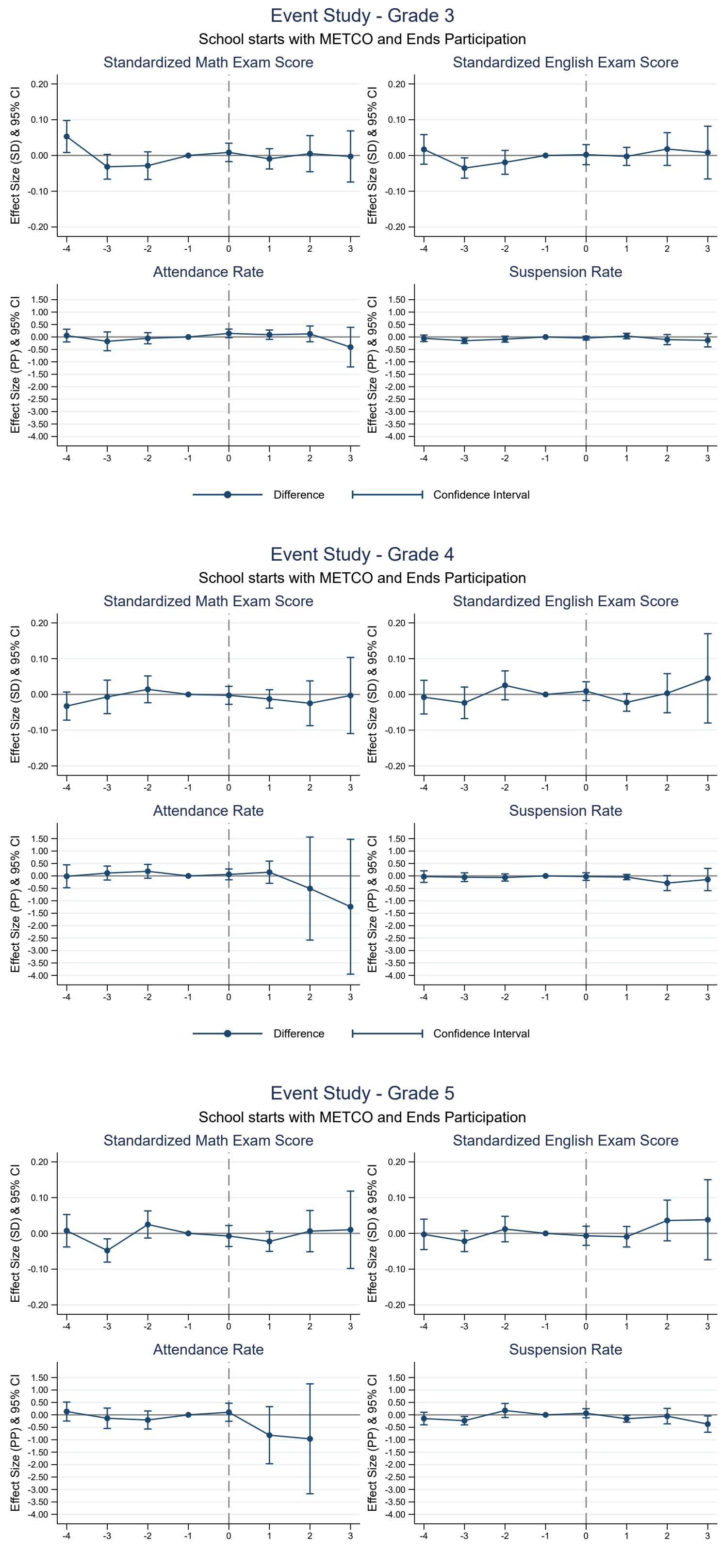
Notes: In Panel A, each bar represents an individual school district and the proportion of its elementary schools that enroll METCO students. In Panel B, each bar represents an individual school district and the proportion of its elementary school cohorts that enroll METCO students. The bars in Panel C represent individual school districts' average proportion of METCO students in a grade for the school cohorts that have any METCO students. All figures only display suburban districts that accept METCO students and have elementary schools. Four districts start METCO enrollment after elementary school.

Figure 2: Event Study Results: Begins METCO Participation



Notes: This figure shows the event study estimates of Equation 4 for grades 3 - 5 outcomes for schools that began their METCO participation.

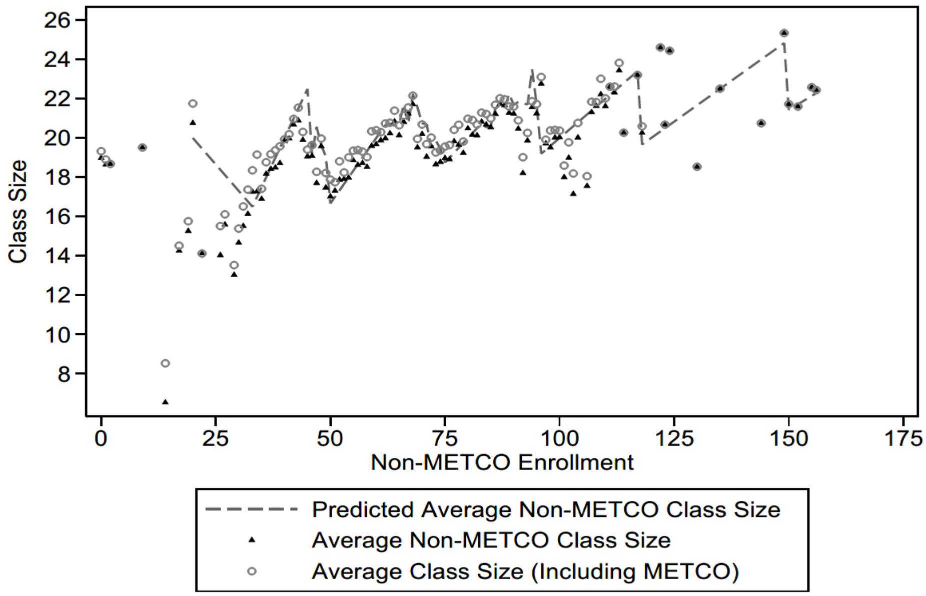
Figure 3: Event Study Results: Ends METCO Participation



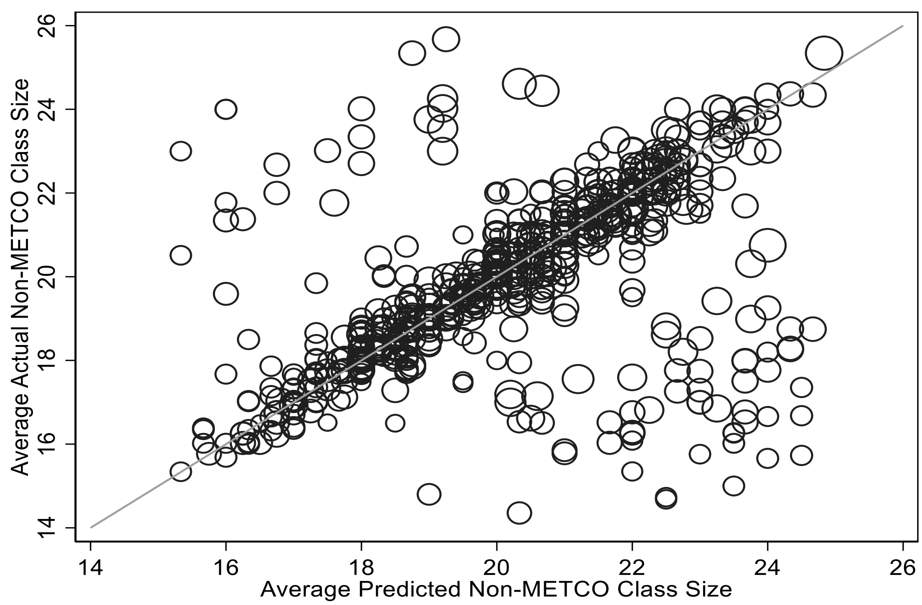
Notes: This figure shows the event study estimates of Equation 4 for grades 3 - 5 outcomes for schools that ended or paused their METCO participation.

Figure 4: Relationship between predicted and actual class size

Panel A : Predicted and actual class size by total non-METCO first grade enrollment

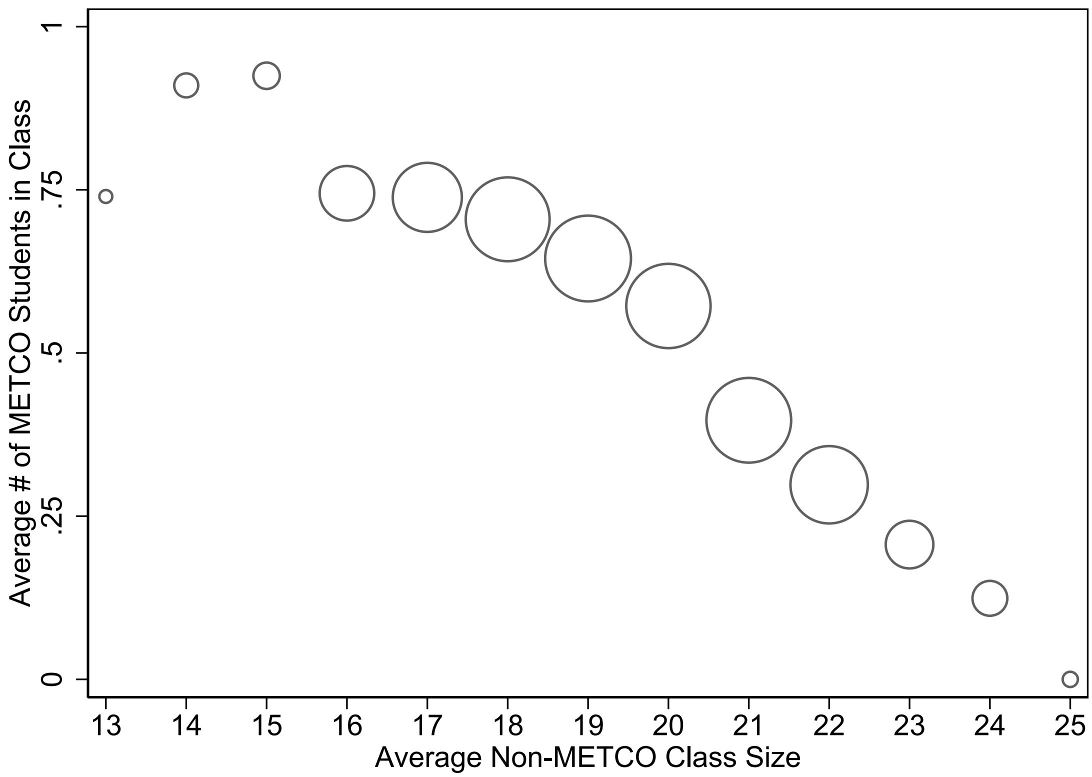


Panel B: Relationship between predicted and actual class size



Notes: In Panel A, each data point reflects an individual school by year for first grade students. Predicted Non-METCO class size is estimated using Angrist and Lavy (1999)'s Maimonides rule with a class size cutoff of 25 students. Actual class size cutoffs in METCO suburban districts range from 20 to 30 students with an average of 24 students per class. Panel B plots a 45 degree line and the average non-METCO class size in suburban METCO receiving districts against the predicted non-METCO class size according to the Maimonides Rule described in Equation 5. Data is for first grades and at the school by year level. The predicted class size calculation uses district-specific class size cutoffs for first grade. The sample includes the 19 districts in the instrumental variable analysis sample.

Figure 5: Relationship between number of METCO students in class and average non-METCO class size



Notes: This figure plots average number of METCO students per class against binned, integer values of average non-METCO class size for the first grade. Circle size is average weighted by first grade enrollment with larger circles signifying more common average non-METCO class sizes. The sample includes the 19 districts in the instrumental variable analysis sample.

Table 1: Descriptive Statistics of K-12 Data

	METCO Suburban Districts						
	Resident Students					Mean	
	Mean	District Level Min	District Level Max	Black & Latinx Residents	Urban Students	Boston & Springfield Public Schools	Non-METCO Districts in Metro-area
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Female	0.49	0.44	0.54	0.49	0.55	0.48	0.49
Black	0.02	0.00	0.09	0.41	0.77	0.36	0.08
Latino/a	0.03	0.00	0.15	0.59	0.17	0.42	0.11
Asian	0.08	0.00	0.41	0.00	0.04	0.07	0.05
White	0.85	0.48	0.99	0.02	0.02	0.14	0.75
Subsidized Lunch	0.06	0.00	0.25	0.36	0.51	0.78	0.24
Special Education	0.13	0.07	0.27	0.17	0.11	0.16	0.14
English Learner	0.05	0.00	0.22	0.18	0.05	0.29	0.10
Immigrant	0.03	0.00	0.19	0.08	0.02	0.12	0.05
Ever Suspended	0.02	0.00	0.12	0.06	0.08	0.10	0.06
Attendance Rate	0.95	0.85	0.98	0.92	0.94	0.87	0.93
Earliest Test Score							
English	0.43	0.02	0.80	-0.13	-0.22	-0.73	0.02
Math	0.45	0.03	0.89	-0.19	-0.34	-0.64	0.03
Infant Weight (in grams)	3433	3282	3527	3306	3245	3240	3408
Married parents	0.93	0.76	1.00	0.62	0.45	0.37	0.80
Absent father at birth	0.02	0.00	0.06	0.15	0.24	0.31	0.06
On Medicaid at birth	0.06	0.00	0.25	0.38	0.38	0.64	0.20
Mother's highest education level							
Less than high school	0.02	0.00	0.08	0.14	0.09	0.34	0.09
High school grad	0.13	0.01	0.37	0.27	0.31	0.38	0.28
Some college	0.15	0.02	0.30	0.19	0.31	0.16	0.20
2 year college	0.05	0.00	0.16	0.05	0.08	0.03	0.07
College or more	0.65	0.23	0.93	0.34	0.20	0.09	0.36
Father's highest education level							
Less than high school	0.02	0.00	0.07	0.11	0.07	0.27	0.08
High school grad	0.16	0.02	0.43	0.30	0.42	0.44	0.34
Some college	0.13	0.01	0.23	0.17	0.25	0.14	0.16
2 year college	0.03	0.00	0.13	0.05	0.06	0.03	0.05
College or more	0.66	0.26	0.94	0.38	0.20	0.12	0.37
Observations with Test	1,241,953			42,615	29,581	460,264	3,337,192
Observations	1,686,366			91,790	47,193	1,107,318	4,607,202
Unique Students	217,135			18,110	6,620	174,127	608,258
Years of Data	19			19	19	19	19
Unique Schools	243			242	198	231	688
Unique Districts	38			38	38	2	101

Notes: This table shows descriptive baseline statistics for students across school types and student characteristics. Data is at the student by year level. Data include Kindergarten through 12th grade for the 2001-2002 through 2019-2020 school years. Test scores are centered at the state mean for each grade and year. State exams are given in grades 3 - 8 and 10. The earliest test score is the first state standardized exam the student took, which is third grade for most students. Parental education, infant weight, family structure, and Medicaid come from Massachusetts birth records data and are only available for those born in Massachusetts. Columns 2 and 3 show the minimum and maximum district average of resident students for each characteristic.

Table 2: Ordinary Least Squares Estimates of the Relationship between METCO and Outcomes

	Metro-Area			Receiving Districts				
	Non-METCO mean	Relationship between METCO and Outcome		Non-METCO mean	Relationship between METCO and Outcome			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Math	0.091	0.408***	0.017***	0.392	0.096***	0.015***	0.004	0.004
		(0.002)	(0.003)		(0.003)	(0.003)	(0.003)	(0.003)
N		3,789,167	3,789,167		1,036,030	1,036,030	957,542	957,541
English	0.099	0.370***	0.013***	0.408	0.074***	0.012***	0.006**	0.006**
		(0.002)	(0.003)		(0.002)	(0.003)	(0.003)	(0.003)
N		3,774,366	3,774,366		1,032,576	1,032,576	954,001	954,000
Attendance Rate	0.927	0.014***	0.004***	0.945	0.000	0.003***	0.002***	0.002***
		(0.000)	(0.000)		(0.000)	(0.000)	(0.000)	(0.000)
N		8,248,960	8,248,960		2,227,579	2,227,579	1,936,005	1,935,993
Suspended	0.047	-0.030***	-0.005***	0.009	-0.004***	0.001***	0.001***	0.001***
		(0.000)	(0.000)		(0.000)	(0.000)	(0.000)	(0.000)
N		7,784,916	7,784,916		2,103,547	2,103,547	1,936,005	1,935,993
Number of Suspensions	0.148	-0.107***	-0.022***	0.019	-0.014***	0.002**	0.003***	0.003***
		(0.001)	(0.001)		(0.001)	(0.001)	(0.001)	(0.001)
N		7,784,916	7,784,916		2,103,547	2,103,547	1,936,005	1,935,993
Year & Grade FE, Region Control		X	X		X	X	X	X
District FE			X			X	X	X
Lagged School Traits							X	X
Individual Baseline Controls								X

Notes: This table reports the OLS estimates of the relationship between the proportion of METCO students in a schoolXgrade and the outcomes of non-METCO students in that suburban district. The endogenous variable is the ratio of METCO to non-METCO students in a schoolXgrade multiplied by 20 so that a one unit increase in the endogenous variable can be interpreted as adding one METCO student to a 20-person class. All models control for whether the district is in the Springfield metropolitan area and include year and grade fixed effects. Lagged school traits include the prior year's average attendance rate, suspension rate, days suspended, and standardized Math and English test scores as well as the proportion of students receiving free or reduced price lunch, special education or English Language Learner services, and racial demographics. Individual baseline controls include the following information for the first year students appear in the data: gender, race and ethnicity, free and reduced price lunch, special education and level of services, English Language Learner and level of proficiency, and immigrant.

Table 3: Difference-in-Differences Results

	School does not start with METCO and Begins Participation			School starts with METCO and Ends Participation		
	Grade 3	Grade 4	Grade 5	Grade 3	Grade 4	Grade 5
	(1)	(2)	(3)	(4)	(5)	(6)
Math	0.002	0.005	-0.004	-0.003	-0.008	-0.016
	(0.015)	(0.018)	(0.016)	(0.014)	(0.013)	(0.015)
Observations	990,108	930,356	481,110	1,174,017	1,081,556	787,377
Treatment Groups	143	144	128	156	163	146
Comparison Groups	4,644	4,685	3,893	5,856	6,107	5,251
English	0.006	0.025*	-0.014	0.008	-0.016	-0.004
	(0.015)	(0.015)	(0.016)	(0.013)	(0.014)	(0.015)
Observations	1,103,777	932,425	481,840	1,283,026	1,073,060	788,929
Treatment Groups	155	144	128	168	163	146
Comparison Groups	5,092	4,685	3,894	6,333	6,107	5,251
Attendance Rate	0.0004	-0.0021	0.0017	0.0000	-0.0022	-0.0086*
	(0.0009)	(0.0043)	(0.0051)	(0.0008)	(0.0032)	(0.0049)
Observations	1,125,637	1,024,827	545,195	1,306,911	1,137,860	875,253
Treatment Groups	155	155	140	168	168	157
Comparison Groups	5,094	5,031	4,284	6,337	6,289	5,632
Suspended	0.0002	0.0011	0.0015	0.0004	-0.0009	-0.0011
	(0.0006)	(0.0007)	(0.0010)	(0.0006)	(0.0007)	(0.0008)
Observations	1,125,566	1,024,805	545,195	1,306,906	1,137,859	875,253
Treatment Groups	155	155	140	168	168	157
Comparison Groups	5,094	5,031	4,284	6,337	6,289	5,632
Number of Suspensions	-0.0006	0.0015	0.0018	0.0023	0.0006	-0.0018
	(0.0014)	(0.0013)	(0.0013)	(0.0015)	(0.0016)	(0.0012)
Observations	1,125,566	1,024,805	545,195	1,306,906	1,137,859	875,253
Treatment Groups	155	155	140	168	168	157
Comparison Groups	5,094	5,031	4,284	6,337	6,289	5,632

Notes: This table shows the stacked difference in differences estimates of the impact of schools switching to enrolling METCO students (Columns 1 – 3) and schools switching to not enrolling METCO students (Columns 4 – 6). The stacked difference in differences calculates a weighted average of the individual difference in differences estimates of each school that changed METCO enrollment policy. Data includes up to four years before the school switched to enrolling METCO first graders and up to four years after. The comparison schools did not accept METCO students in first grade for all the years the treatment school appears in the data. Each individual treatment group and their comparison schools form a policy change subsample. The treatment and comparison schools are chosen using the same rules for the endogenous variable of switching to no longer enrolling METCO students in Columns 4–6. Observations can appear in multiple comparison groups for treatment schools. Treatment observations can appear as comparison observations for other treatment schools. Standard errors are clustered at the individual student level and by the policy change subsample. All estimates include school fixed effects and individual student demographics (gender, race, immigrant status). The sample includes students who entered the first grade in the 2001-2002 through 2017-2018 school years and excludes METCO participants. Attendance and suspension estimates for grade one and two are similarly small, close to zero and not statistically significant.

Table 4: Two-Stage Least Squares Results

	3rd Grade Mean	Grade 1	Grade 2	Grade 3	Grade 4	Grade 5
	(1)	(2)	(3)	(4)	(5)	(6)
Math	0.480			0.020	0.036	0.009
				(0.035)	(0.035)	(0.031)
N				58,907	58,705	58,528
F-stat				225	277	395
English	0.492			0.049	0.067	0.016
				(0.034)	(0.033)	(0.029)
N				58,907	58,705	58,528
F-stat				225	277	395
Attendance Rate	0.967	-0.003	-0.011	0.006	0.000	0.005
		(0.004)	(0.002)	(0.002)	(0.002)	(0.001)
N		59,720	59,138	58,907	58,705	58,528
F-stat		142	164	225	277	395
Suspended	0.002	-0.002	-0.002	-0.002	-0.001	-0.001
		(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
N		59,650	59,138	58,907	58,705	58,528
F-stat		142	164	225	277	395
Number of Discrete Class Sizes		10	10	10	10	10
Number of Cohorts (SchoolXYear)		1053	1084	1085	1085	1084
Number of Schools		90	91	91	91	91
Number of Districts		19	19	19	19	19

Notes: This table shows the two-stage least squares estimates of the impact of an additional METCO student on suburban resident student outcomes. The main endogenous variable is 20 times the average ratio of METCO to non-METCO students in an individual's grade cohort from grade 1 through the outcome grade. A one unit increase in the main endogenous variable represents an additional METCO student in a 20-student classroom. The model includes another endogenous variable: average class size excluding METCO students. This is estimated for 2001-2011 before administrative data on class size exists (see Equation 9). Linear and integer forms of predicted class size (see Equation 6 for estimation strategy) instrument for both the ratio of METCO to non-METCO students and non-METCO class size. Controls include binned total grade-level enrollment that proxy for the number of classrooms in a grade, school and year fixed effects, individual baseline covariates (including gender and race). The sample excludes METCO participants and includes student with non-missing values for the third through fifth grade Math and English exams. Those students entered first grade between 2003-2004 through 2014-15 school year. The included districts each have significant first stage estimates with F-statistics above 100 for grades 3 through 5.

Table 5: Estimates for School Switching

	School does not start with METCO and Begins Participation				School starts with METCO and Ends Participation				Instrumental Variable Estimates			
	Switch Schools within District		Leave School District		Switch Schools within District		Leave School District		Switch Schools within District		Leave School District	
	Non-METCO Mean, Pre-period	Effect	Non-METCO Mean, Pre-period	Effect	Non-METCO Mean, Pre-period	Effect	Non-METCO Mean, Pre-period	Effect	Mean	Effect	Mean	Effect
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Switch in Grade 1	0.003	0.000	0.002	0.000	0.002	-0.001**	0.002	-0.001***	0.002	0.000	0.001	0.001
		(0.000)		(0.000)		(0.000)		(0.000)		(0.003)		(0.002)
Switch by Grade 2	0.050	-0.002	0.057	0.006**	0.033	-0.001	0.064	-0.006**	0.039	0.004	0.057	0.004
		(0.004)		(0.002)		(0.004)		(0.002)		(0.010)		(0.006)
Switch by Grade 3	0.081	0.005	0.099	0.004	0.064	-0.011	0.105	-0.002	0.066	-0.025**	0.099	0.000
		(0.009)		(0.003)		(0.009)		(0.003)		(0.011)		(0.007)
Switch by Grade 4	0.106	0.005	0.124	0.005	0.089	-0.015*	0.140	-0.003	0.093	-0.013	0.132	0.011
		(0.010)		(0.003)		(0.009)		(0.003)		(0.016)		(0.010)
Switch by Grade 5	0.141	0.014	0.137	0.007**	0.101	-0.024**	0.165	-0.004	0.107	-0.020	0.153	-0.012
		(0.014)		(0.004)		(0.010)		(0.004)		(0.012)		(0.008)
N	584125	1411422	584125	1411422	1178041	1901351	1178041	1901351	122208	85580	122208	85580
# Treatment Groups		171		171		211		211				
# Control Schools		5632		5632		7727		7727				

Notes: This table shows the stacked difference in differences and two-stage least squares estimates of the impact of having METCO students in the grade cohort on switching schools after first grade.Standard errors are clustered at the individual student level and by the policy change subsample for the difference in differenes models. Difference in differences estimates include school fixed effects and individual student demographics (gender, race, immigrant status). The sample excludes METCO participants. See Tables 3 and 4 for full specifications. Students who switch to private or out of state schools are counted as leaving the district. Non-voluntary school moves due to school closures or reaching the highest grade offerred in a school are not counted as school switching.

Table 6: Estimates for Class and Teacher Traits

	School does not start with METCO and Begins			School starts with METCO and Ends Participation			Instrumental Variable Analysis		
	Untreated mean	Effect of having METCO	SE	Untreated mean	Effect of not having METCO	SE	Mean	Effect	SE
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Panel A: Teacher Traits									
% of Teachers licensed in teaching assignment	1.000	0.001	(0.001)	0.999	0.002***	(0.000)	1.000	0.000	(0.001)
% of core academic classes taught by highly	0.995	-0.001	(0.003)	0.999	0.001	(0.002)	0.998	0.002	(0.003)
% of teachers with advanced degree	0.020	0.001	(0.011)	0.011	-0.003	(0.009)	0.012	0.000	(0.007)
Average years of teaching experience in MA	10.805	0.361	(0.292)	10.875	-0.157	(0.285)	10.795	-1.002***	(0.334)
% novice teacher (<2 years)	0.102	-0.009	(0.019)	0.091	0.018	(0.015)	0.099	0.022	(0.014)
Any white teacher	0.996	-0.003	(0.007)	0.938	0.003	(0.007)	0.958	0.026**	(0.010)
Any Hispanic teacher	0.014	-0.003	(0.003)	0.020	-0.006**	(0.003)	0.010	-0.002	(0.005)
Any Black teacher	0.001	0.001	(0.004)	0.023	-0.004	(0.003)	0.016	-0.011*	(0.006)
Any Asian teacher	0.003	0.005	(0.007)	0.042	-0.004	(0.005)	0.028	-0.036***	(0.008)
Panel B: Class Traits									
Class 90th - 10th percentile (all students)									
Math	1.953	0.009	(0.035)	1.952	-0.022	(0.032)	1.886	0.123***	(0.027)
English	1.894	0.038	(0.031)	1.883	-0.042	(0.033)	1.831	0.184***	(0.027)
Average # METCO students	0.029	0.747***	(0.048)	0.929	-0.730***	(0.042)	0.635	1.032***	(0.039)
Average % special education	0.188	-0.007	(0.005)	0.184	0.007	(0.004)	0.174	-0.018***	(0.006)
Average % free-reduced lunch	0.113	0.014***	(0.004)	0.090	-0.014***	(0.004)	0.091	0.017***	(0.004)
Average % English Learner	0.032	0.003	(0.002)	0.059	-0.002	(0.003)	0.043	-0.003	(0.004)
Average class size	20.794	-0.678***	(0.194)	21.435	0.452**	(0.214)	22.210	-2.476***	(0.127)
Suspension rate	0.002	0.001	(0.001)	0.006	-0.001	(0.001)	0.004	-0.001	(0.001)
Residents' suspension rate	0.002	0.001	(0.001)	0.005	0.000	(0.001)	0.004	0.000	(0.001)
Average attendance rate	0.955	-0.003	(0.006)	0.956	-0.005	(0.005)	0.962	0.000	(0.001)
Average attendance of residents	0.955	-0.003	(0.006)	0.956	-0.005	(0.005)	0.962	-0.001	(0.001)

Notes: Column 1 shows the average classroom traits for core classes without METCO students using class by year level data. Columns 2 and 3 show stacked difference in differences estimates of the impact of schools switching to enrolling METCO students on fourth grade teacher and classroom traits. Columns 5 and 6 show the estimates for schools switching to not enrolling METCO students. Columns 8 and 9 show the two-stage least squares estimates. Only core academic classes are included. See Tables 3 and 4 for regression specifications.

Table 7: Two-Stage Least Squares High School Outcomes Results

	Non-METCO mean (1)	2SLS Estimate (2)	Standard Error (3)	N (4)	F-stat (5)
Panel A: Testing					
Meets Standardized Testing High School Graduation Requirement	0.939	0.007	(0.005)	43765	796
Qualify for Adams Scholarship	0.297	0.011	(0.010)	43814	796
Take SAT	0.636	0.028***	(0.010)	51864	869
SAT 800 or Higher	0.623	0.032***	(0.010)	51864	869
SAT 1000 or Higher	0.539	0.035***	(0.011)	51864	869
SAT 1200 or Higher	0.328	0.010	(0.010)	51864	869
SAT 1400 or Higher	0.110	0.015**	(0.007)	51864	869
Took AP	0.490	0.014	(0.011)	51864	869
Number of APs	1.650	0.121**	(0.048)	51864	869
AP Score Above 3	0.438	0.014	(0.011)	51864	869
AP Score 4 or 5	0.358	0.021**	(0.010)	51864	869
Panel B: High School Graduation					
Dropout	0.005	0.000	(0.002)	53649	815
Graduate in 4 Years	0.874	0.000	(0.005)	53649	815
Graduate in 5 Years	0.981	0.002	(0.003)	48127	739
Panel C: Post High School Aspirations					
Any College	0.912	0.003	(0.007)	43715	751
2 Year College	0.047	0.006	(0.005)	43715	751
4 Year College	0.865	-0.003	(0.008)	43715	751
Number of Discrete Class Sizes		11			
Number of Cohorts (SchoolXYear)		878			
Number of Schools		93			
Number of Districts		19			

Notes: This table shows the two-stage least squares estimates of the impact of an additional METCO student on suburban resident student high school outcomes. The main endogenous variable is 20 times the average ratio of METCO to non-METCO students in an individual's grade cohort from grade 1. A one unit increase in the main endogenous variable represents an additional METCO student in a 20-student classroom. The model includes another endogenous variable: average class size excluding METCO students. This is estimated for 2001-2011 before administrative data on class size exists (see Equation 9). Linear and integer forms of predicted class size (see Equation 6 for estimation strategy) instrument for both the ratio of METCO to non-METCO students and non-METCO class size. Controls include binned total grade-level enrollment that proxy for the number of classrooms in a grade, school and year fixed effects, individual baseline covariates (including gender, race, free and reduced price lunch status, special education status, and English Language Learner status). Students have a value of "0" for SAT and AP outcomes if they did not take the respective exams. Post-high school plans data come from a survey administered to 10th graders.

Table 8: Two-Stage Least Squares College Outcomes Results

	Non-METCO mean (1)	2SLS Estimate (2)	Standard Error (3)	N (4)	F-stat (5)
Panel A: Racial Composition of College					
Percent Black or Latinx	0.155	-0.003	(0.002)	34261	662
Percent Non-white	0.383	-0.001	(0.004)		
Panel B: College Enrollment					
Any college	0.730	0.018**	(0.008)	51841	868
Two-Year college	0.077	-0.005	(0.006)		
Four-year college	0.691	0.024***	(0.009)		
Four-year Public	0.275	-0.002	(0.010)		
Four-year Private	0.444	0.029***	(0.010)		
Four-year Public in MA	0.134	-0.001	(0.007)		
Panel C: Four-Year College Ranking					
Most Competitive	0.139	0.018**	(0.007)	51841	868
Highly Competitive	0.276	0.030***	(0.009)		
Very Competitive	0.401	0.027***	(0.010)		
Competitive	0.660	0.018**	(0.009)		
Panel D: Four-Year College Persistence					
One academic semester	0.513	0.026**	(0.010)	51841	868
Three academic semesters	0.503	0.023**	(0.010)	46464	860
Five academic semesters	0.495	0.031***	(0.011)	41192	778
Seven academic semesters	0.487	0.028**	(0.011)	35956	801
Panel E: College Graduation within 6 years					
Any	0.662	0.028**	(0.012)	30505	723
Two-Year college	0.022	-0.003	(0.004)		
Four-year college	0.649	0.031**	(0.012)		

Notes: This table shows the two-stage least squares estimates of the impact of an additional METCO student on suburban resident student college outcomes. The main endogenous variable is 20 times the average ratio of METCO to non-METCO students in an individual's grade cohort from grade 1. A one unit increase in the main endogenous variable represents an additional METCO student in a 20-student classroom. The model includes another endogenous variable: average class size excluding METCO students. This is estimated for 2001-2011 before administrative data on class size exists (see Equation 9). Linear and integer forms of predicted class size (see Equation 6 for estimation strategy) instrument for both the ratio of METCO to non-METCO students and non-METCO class size. Controls include binned total grade-level enrollment that proxy for the number of classrooms in a grade, school and year fixed effects, individual baseline covariates (including gender, race, free and reduced price lunch status, special education status, and English Language Learner status). Racial composition of college equals 0 if the student does not enroll in college and is for the first college the student enrolled in. The racial composition data comes from IPEDS annual data. Four-year college enrollment includes students who start in two-year colleges and transfer. Competitiveness rankings are determined by Barron's for the first four-year college the student enrolled in (if any). Panel B measures college enrollment within 18 months of the student's projected 4-year high school graduation. Panel D measures college persistence for on-time college enrollment (within 6 months of the student's projected 4-year high school graduation date).

Table 9: Differences between Classes with and without METCO Students in Treated Cohorts

	Average for class with no METCO students	Difference between class with METCO and no METCO	
		Coefficient	SE
	(1)	(2)	(3)
Panel A: Teacher Traits			
% of Teachers licensed in teaching assignment	0.998	0.000	(0.000)
% of core academic classes taught by highly qualified	0.997	0.000	(0.001)
% of teachers with advanced degree	0.029	0.004***	(0.002)
Average years of teaching experience in MA	12.002	0.203***	(0.078)
% novice teacher (<2 years)	0.097	-0.010***	(0.003)
% new to school (<2 years in school)	0.063	-0.007***	(0.002)
Any white teacher	0.969	-0.007***	(0.002)
Any Hispanic teacher	0.009	0.003***	(0.001)
Any Black teacher	0.012	0.010***	(0.001)
Any Asian teacher	0.024	0.002	(0.002)
Any Non-white teachers	0.046	0.014***	(0.002)
Teacher Value Added for Non-METCO students - Math	-0.044	0.144***	(0.029)
Teacher Value Added for Non-METCO students - English	0.171	-0.037	(0.029)
Panel B: Class Traits			
Average lagged test score of residents:			
Math	0.418	-0.003	(0.005)
English	0.420	0.003	(0.005)
Class 90th - 10th percentile (all students)			
Math	1.905	0.174***	(0.009)
English	1.849	0.137***	(0.009)
Average # METCO students	0.000	1.441***	(0.005)
Average % special education	0.179	-0.005***	(0.001)
Average % free-reduced lunch	0.098	0.029***	(0.001)
Average % English Learner	0.069	-0.004***	(0.001)
Average class size	19.617	0.552***	(0.033)
Suspension rate	0.004	0.001***	(0.000)
Residents' suspension rate	0.004	0.000	(0.000)
Lagged suspension rate	0.003	0.000	(0.000)
Lagged residents' suspension rate	0.003	0.000*	(0.000)
Average attendance rate	0.891	0.002***	(0.001)
Average attendance of residents	0.891	0.003***	(0.001)
Lagged attendance rate	0.946	-0.001*	(0.001)
Average lagged attendance rate of residents	0.946	-0.001	(0.001)
Average % Black	0.019	0.049***	(0.001)
Average % Black among residents	0.019	0.003***	(0.000)
Average % Latinx	0.057	0.013***	(0.001)
Average % Latinx among residents	0.057	-0.002**	(0.001)
N	19504		

Notes: This table shows the average traits of classroom and teacher characteristics for core-subject classes without METCO students in column 1. Column 2 displays the relationship between having any METCO students in a class and the class' characteristics after controlling for year, school, and grade indicators. Lagged test score data means and distributions only include 4th and 5th grade classrooms because the test is first administed in 3rd grade. Other rows include first through fifth grade classrooms.